

# Fundamental Results on Fluid Approximations of Stochastic Process Algebra Models

Jie Ding

*School of Information Engineering, The University of Yangzhou, Yangzhou, China*

Jane Hillston

*Laboratory for Foundations of Computer Science, School of Informatics, The University of Edinburgh, Edinburgh, UK.*

---

## Abstract

In order to avoid the state space explosion problem encountered in the quantitative analysis of large scale PEPA models, a fluid approximation approach has recently been proposed, which results in a set of ordinary differential equations (ODEs) to approximate the underlying continuous time Markov chain (CTMC). This paper presents a mapping semantics from PEPA to ODEs based on a numerical representation scheme, which extends the class of PEPA models that can be subjected to fluid approximation. Furthermore, we have established the fundamental characteristics of the derived ODEs, such as the existence, uniqueness, boundedness and nonnegativeness of the solution. The convergence of the solution as time tends to infinity for several classes of PEPA models, has been proved under some mild conditions. For general PEPA models, the convergence is proved under a particular condition, which has been revealed to relate to some famous constants of Markov chains such as the spectral gap and the Log-Sobolev constant. This thesis has established the consistency between the fluid approximation and the underlying CTMCs for PEPA, i.e. the limit of the solution is consistent with the equilibrium probability distribution corresponding to a family of underlying density dependent CTMCs.

*Keywords:* Fluid Approximation; PEPA; Convergence

---

*Email addresses:* jieding@yzu.edu.cn (Jie Ding), jane.hillston@ed.ac.uk (Jane Hillston)

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>The PEPA modelling formalism</b>	<b>5</b>
2.1	Introduction to PEPA . . . . .	5
2.2	Numerical Representation of PEPA Models . . . . .	7
<b>3</b>	<b>Fluid approximations for PEPA models</b>	<b>11</b>
3.1	State space explosion problem: an illustration by a tiny example	11
3.2	Fluid approximation of PEPA models . . . . .	13
3.3	Existence and uniqueness of ODEs' solution . . . . .	15
3.4	Convergence and consistence of ODEs' solution: nonsynchro- nised models . . . . .	16
3.4.1	Features of ODEs without synchronisations . . . . .	16
3.4.2	Convergence and consistency for the ODEs . . . . .	19
<b>4</b>	<b>Relating to density dependent CTMCs</b>	<b>20</b>
4.1	Density dependent Markov chains underlying PEPA models .	21
4.2	Fluid approximation as the limit of the CTMCs . . . . .	24
4.3	Consistency between the fluid approximation and the CTMCs	25
<b>5</b>	<b>Convergence of ODEs' solution: a probabilistic approach</b>	<b>26</b>
5.1	Convergence under a particular condition . . . . .	26
5.2	Investigation of the particular condition . . . . .	30
5.2.1	Important estimation for Markov kernel . . . . .	30
5.2.2	Study of the particular condition . . . . .	32
<b>6</b>	<b>Fluid analysis: an analytic approach (I)</b>	<b>35</b>
6.1	Analytical proof of boundedness and nonnegativeness . . . . .	35
6.2	A case study on convergence with two synchronisations . . . . .	38
6.2.1	Theoretical study of convergence . . . . .	38
6.2.2	Experimental study of convergence . . . . .	47
<b>7</b>	<b>Fluid analysis: an analytic approach (II)</b>	<b>48</b>
7.1	Convergence for two component types and one synchronisation (I): a special case . . . . .	48

7.1.1	A previous model and the fluid approximation . . . . .	49
7.1.2	Proof outline of convergence . . . . .	50
7.1.3	Proof not relying on explicit expressions . . . . .	53
7.2	Convergence for two component types and one synchronisation	
	(II): general case . . . . .	60
7.2.1	Features of coefficient matrix . . . . .	60
7.2.2	Eigenvalues of coefficient matrix . . . . .	65
7.2.3	Convergence theorem . . . . .	66
<b>8</b>	<b>Related work</b>	<b>69</b>
<b>9</b>	<b>Conclusions</b>	<b>70</b>
	<b>Appendix A</b>	<b>77</b>
	<b>Appendix B</b>	<b>78</b>
	<b>Appendix C</b>	<b>80</b>
	<b>Appendix D</b>	<b>81</b>
	Appendix.D.1 . . . . .	81
	Appendix.D.2 . . . . .	83
	Appendix.D.3 . . . . .	87

## 1. Introduction

Stochastic process algebras, such as PEPA [1], TIPP [2], EMPA [3], IMC [4], are powerful modelling formalisms for concurrent systems which have enjoyed considerable success over the last decade. Such modeling can help designers and system managers by allowing aspects of a system which are not readily tested, such as scalability, to be analysed before a system is deployed. However, both model construction and analysis can be challenged by the size and complexity of large scale systems. This problem, called *state space explosion*, is inherent in the discrete state approach employed in stochastic process algebras and many other formal modelling approaches. To overcome this problem, many work devoted exploiting the compositionality of the process algebra to decompose or simplify the underlying CTMC e.g. [5, 6, 7, 8, 9, 10]. Another technique is the use of Kronecker algebra [11, 12, 13, 14]. In addition, abstract Markov chains and stochastic

bounds techniques have also been used to analyse large scale PEPA models [15, 16].

The techniques reported above are based on the discrete state space. Therefore as the size of the state space is extremely large, these techniques are not always strong enough to handle the state space explosion problem. For example, in the modelling of biochemical mechanisms using the stochastic  $\pi$ -calculus [17, 18] and PEPA [19, 20], the state space explosion problem becomes almost insurmountable. Consequently in many cases models are analysed by discrete event simulation rather than being able to abstractly consider all possible behaviours.

To avoid this problem Hillston proposed a radically different approach in [21] from the following two perspectives: choosing a more abstract state representation in terms of state variables, quantifying the types of behaviour evident in the model; and assuming that these state variables are subject to continuous rather than discrete change. This approach results in a set of ODEs, leading to the evaluation of transient, and in the limit, steady state measures.

However, there are not many discussions on the fundamental problems, such as the existence, uniqueness, boundedness and nonnegativeness of the solution, as well as the its asymptotic behaviour as time tends to infinity, and the relationship between the derived ODEs and the underlying CTMCs for general PEPA models. Solving these problems can not only bring confidence in the new approach, but can also provide new insight into, as well as a profound understanding of, performance formalisms. This paper will focus on these topics and give answers to these problems.

The remainder of this paper is structured as follows. Section 2 will give a brief introduction to PEPA as well as a numerical representation scheme developed for PEPA. Based on this scheme, the fluid approximation of PEPA models will be introduced in Section 3. In this section, the existence and uniqueness of solutions of the derived ODEs will be presented. In addition, we will show that for a PEPA model without synchronisations, the solution of the ODEs converges as time goes to infinity and the limit coincides with the steady-state probability distribution of the underlying CTMC. In Section 4, we demonstrate the consistency between the fluid approximation and the Markov chains underlying the same PEPA model. This relationship will be utilised to investigate the long-time behaviour of the ODEs' solutions in Section 5. The convergence of the solutions will be proved under a particular condition, which relates the convergence problem to some well-known

constants of Markov chains such as the spectral gap and the Log-Sobolev constant. Section 6 and 7 present an analytic approach to analyse the fluid approximation. For several classes of PEPA models, the convergence will be demonstrated under some mild conditions, and the coefficient matrices of the derived ODEs have been exposed to have the following property: all eigenvalues are either zeros or have negative real parts. In addition, the structural property of invariance in PEPA models will be shown to play an important role in the proof of convergence. Finally, after presenting some related work in Section 8, we conclude the paper in Section 9.

## 2. The PEPA modelling formalism

This section will briefly introduce the PEPA language and its numerical representation scheme. The numerical representation scheme for PEPA was developed by Ding in his thesis [22], and represents a model numerically rather than syntactically supporting the use of mathematical tools and methods to analyse the model.

### 2.1. Introduction to PEPA

PEPA (Performance Evaluation Process Algebra) [1], developed by Hillston in the 1990s, is a high-level model specification language for low-level stochastic models, and describes a system as an interaction of the components which engage in activities. In contrast to classical process algebras, activities are assumed to have a duration which is a random variable governed by an exponential distribution. Thus each activity in PEPA is a pair  $(\alpha, r)$  where  $\alpha$  is the action type and  $r$  is the activity rate. The language has a small number of combinators, for which we provide a brief introduction below; the structured operational semantics can be found in [1]. The grammar is as follows:

$$\begin{aligned} S &::= (\alpha, r).S \mid S + S \mid C_S \\ P &::= P \boxtimes_L P \mid P/L \mid C \end{aligned}$$

where  $S$  denotes a *sequential component* and  $P$  denotes a *model component* which executes in parallel.  $C$  stands for a constant which denotes either a sequential component or a model component as introduced by a definition.  $C_S$  stands for constants which denote sequential components. The effect of this syntactic separation between these types of constants is to constrain legal PEPA components to be cooperations of sequential processes.

**Prefix:** The prefix component  $(\alpha, r).S$  has a designated first activity  $(\alpha, r)$ , which has action type  $\alpha$  and a duration which satisfies exponential distribution with parameter  $r$ , and subsequently behaves as  $S$ .

**Choice:** The component  $S + T$  represents a system which may behave either as  $S$  or as  $T$ . The activities of both  $S$  and  $T$  are enabled. Since each has an associated rate there is a *race condition* between them and the first to complete is selected. This gives rise to an implicit probabilistic choice between actions dependent of the relative values of their rates.

**Hiding:** Hiding provides type abstraction, but note that the duration of the activity is unaffected. In  $P/L$  all activities whose action types are in  $L$  appear as the “private” type  $\tau$ .

**Cooperation:**  $P \bowtie_L Q$  denotes cooperation between  $P$  and  $Q$  over action types in the cooperation set  $L$ . The cooperands are forced to synchronise on action types in  $L$  while they can proceed independently and concurrently with other enabled activities (*individual* activities). The rate of the synchronised or *shared* activity is determined by the slower cooperation (see [1] for details). We write  $P \parallel Q$  as an abbreviation for  $P \bowtie_{\emptyset} Q$  when  $L = \emptyset$  and  $P[N]$  is used to represent  $N$  copies of  $P$  in a parallel, i.e.  $P[3] = P \parallel P \parallel P$ .

**Constant:** The meaning of a constant is given by a defining equation such as  $A \stackrel{\text{def}}{=} P$ . This allows infinite behaviour over finite states to be defined via mutually recursive definitions.

On the basis of the operational semantic rules (please refer to [1] for details), a PEPA model may be regarded as a labelled multi-transition system

$$\left( \mathcal{C}, \mathcal{Act}, \left\{ \xrightarrow{(\alpha, r)} \mid (\alpha, r) \in \mathcal{Act} \right\} \right)$$

where  $\mathcal{C}$  is the set of components,  $\mathcal{Act}$  is the set of activities and the multi-relation  $\xrightarrow{(\alpha, r)}$  is given by the rules. If a component  $P$  behaves as  $Q$  after it completes activity  $(\alpha, r)$ , then denote the transition as  $P \xrightarrow{(\alpha, r)} Q$ .

The memoryless property of the exponential distribution, which is satisfied by the durations of all activities, means that the stochastic process underlying the labelled transition system has the Markov property. Hence the underlying stochastic process is a CTMC. Note that in this representation the states of the system are the syntactic terms derived by the operational semantics. Once constructed the CTMC can be used to find steady state or transient probability distributions from which quantitative performance can be derived.

## 2.2. Numerical Representation of PEPA Models

As explained above there have been two key steps in the use of fluid approximation for PEPA models: firstly, the shift to a numerical vector representation of the model, and secondly, the use of ordinary differential equations to approximate the dynamic behaviour of the underlying CTMC. In this paper we are only concerned with the former modification.

This section presents the numerical representation of PEPA models developed in [22]. For convenience, we may represent a transition  $P \xrightarrow{(\alpha, r)} Q$  as  $P \xrightarrow{(\alpha, r_\alpha^{P \rightarrow Q})} Q$ , or often simply as  $P \xrightarrow{\alpha} Q$  since the rate is not pertinent to structural analysis, where  $P$  and  $Q$  are two local derivatives. Following [21], hereafter the term *local derivative* refers to the local state of a single sequential component. In the standard structured operational semantics of PEPA used to derive the underlying CTMC, the state representation is syntactic, keeping track of the local derivative of each component in the model. In the alternative *numerical vector representation* some information is lost as states only record the number of instances of each local derivative:

**Definition 1. (Numerical Vector Form [21]).** For an arbitrary PEPA model  $\mathcal{M}$  with  $n$  component types  $C_i, i = 1, 2, \dots, n$ , each with  $d_i$  distinct local derivatives, the numerical vector form of  $\mathcal{M}$ ,  $\mathbf{m}(\mathcal{M})$ , is a vector with  $d = \sum_{i=1}^n d_i$  entries. The entry  $\mathbf{m}[C_{i_j}]$  records how many instances of the  $j$ th local derivative of component type  $C_i$  are exhibited in the current state.

In the following we will find it useful to distinguish derivatives according to whether they enable an activity, or are the result of that activity:

**Definition 2. (Pre and post local derivative)**

1. If a local derivative  $P$  can enable an activity  $\alpha$ , that is  $P \xrightarrow{\alpha} \cdot$ , then  $P$  is called a *pre local derivative* of  $\alpha$ . The set of all pre local derivatives of  $\alpha$  is denoted by  $\text{pre}(\alpha)$ , called the *pre set* of  $\alpha$ .
2. If  $Q$  is a local derivative obtained by firing an activity  $\alpha$ , i.e.  $\cdot \xrightarrow{\alpha} Q$ , then  $Q$  is called a *post local derivative* of  $\alpha$ . The set of all post local derivatives is denoted by  $\text{post}(\alpha)$ , called the *post set* of  $\alpha$ .
3. The set of all the local derivatives derived from  $P$  by firing  $\alpha$ , i.e.

$$\text{post}(P, \alpha) = \{Q \mid P \xrightarrow{\alpha} Q\},$$

is called the post set of  $\alpha$  from  $P$ .

Within a PEPA model there may be many instances of the same activity type but we will wish to identify those that have exactly the same effect within the model. In order to do this we additionally label activities according to the derivatives to which they relate, giving rise to *labelled activities*:

**Definition 3. (*Labelled Activity*).**

1. For any individual activity  $\alpha$ , for each  $P \in \text{pre}(\alpha), Q \in \text{post}(P, \alpha)$ , label  $\alpha$  as  $\alpha^{P \rightarrow Q}$ .
2. For a shared activity  $\alpha$ , for each  $(Q_1, Q_2, \dots, Q_k)$  in

$$\text{post}(\text{pre}(\alpha)[1], \alpha) \times \text{post}(\text{pre}(\alpha)[2], \alpha) \times \dots \times \text{post}(\text{pre}(\alpha)[k], \alpha),$$

label  $\alpha$  as  $\alpha^w$ , where

$$w = (\text{pre}(\alpha)[1] \rightarrow Q_1, \text{pre}(\alpha)[2] \rightarrow Q_2, \dots, \text{pre}(\alpha)[k] \rightarrow Q_k).$$

Each  $\alpha^{P \rightarrow Q}$  or  $\alpha^w$  is called a **labelled activity**. The set of all labelled activities is denoted by  $\mathcal{A}_{\text{label}}$ . For the above labelled activities  $\alpha^{P \rightarrow Q}$  and  $\alpha^w$ , their respective pre and post sets are defined as

$$\text{pre}(\alpha^{P \rightarrow Q}) = \{P\}, \text{post}(\alpha^{P \rightarrow Q}) = \{Q\},$$

$$\text{pre}(\alpha^w) = \text{pre}(\alpha), \text{post}(\alpha^w) = \{Q_1, Q_2, \dots, Q_k\}.$$

In the numerical representation scheme, the transitions between states of the model are represented by a matrix, termed the *activity matrix* — this records the impact of the labelled activities on the local derivatives.

**Definition 4. (*Activity Matrix, Pre Activity Matrix, Post Activity Matrix*).** For a model with  $N_{\mathcal{A}_{\text{label}}}$  labelled activities and  $N_{\mathcal{D}}$  distinct local derivatives, the activity matrix  $\mathbf{C}$  is an  $N_{\mathcal{D}} \times N_{\mathcal{A}_{\text{label}}}$  matrix, and the entries are defined as following

$$\mathbf{C}(P_i, \alpha_j) = \begin{cases} +1 & \text{if } P_i \in \text{post}(\alpha_j) \\ -1 & \text{if } P_i \in \text{pre}(\alpha_j) \\ 0 & \text{otherwise} \end{cases}$$



where  $\alpha_j$  is a labelled activity. The pre activity matrix  $\mathbf{C}^{\text{pre}}$  and post activity matrix  $\mathbf{C}^{\text{post}}$  are defined as

$$\mathbf{C}^{\text{Pre}}(P_i, \alpha_j) = \begin{cases} +1 & \text{if } P_i \in \text{pre}(\alpha_j) \\ 0 & \text{otherwise.} \end{cases},$$

$$\mathbf{C}^{\text{Post}}(P_i, \alpha_j) = \begin{cases} +1 & \text{if } P_i \in \text{post}(\alpha_j) \\ 0 & \text{otherwise.} \end{cases}$$

From Definitions 3 and 4, each column of the activity matrix corresponds to a system transition and each transition can be represented by a column of the activity matrix. The activity matrix equals the difference between the pre and post activity matrices, i.e.  $\mathbf{C} = \mathbf{C}^{\text{Post}} - \mathbf{C}^{\text{Pre}}$ . The rate of the transition between states is specified by a *transition rate function*, but we omit this detail here since we are concerned with qualitative analysis. See [22] for details.

We first give the definition of the apparent rate of an activity in a local derivative.

**Definition 5. (*Apparent Rate of  $\alpha$  in  $P$* )** Suppose  $\alpha$  is an activity of a PEPA model and  $P$  is a local derivative enabling  $\alpha$  (i.e.  $P \in \text{pre}(\alpha)$ ). Let  $\text{post}(P, \alpha)$  be the set of all the local derivatives derived from  $P$  by firing  $\alpha$ , i.e.  $\text{post}(P, \alpha) = \{Q \mid P \xrightarrow{(\alpha, r_\alpha^{P \rightarrow Q})} Q\}$ . Let

$$r_\alpha(P) = \sum_{Q \in \text{post}(P, \alpha)} r_\alpha^{P \rightarrow Q}. \quad (1)$$

The apparent rate of  $\alpha$  in  $P$  in state  $\mathbf{x}$ , denoted by  $r_\alpha(\mathbf{x}, P)$ , is defined as

$$r_\alpha(\mathbf{x}, P) = \mathbf{x}[P]r_\alpha(P). \quad (2)$$

The above definition is used to define the following transition rate function.

**Definition 6. (*Transition Rate Function*)** Suppose  $\alpha$  is an activity of a PEPA model and  $\mathbf{x}$  denotes a state vector.

1. If  $\alpha$  is individual, then for each  $P \xrightarrow{(\alpha, r_\alpha^{P \rightarrow Q})} Q$ , the transition rate function of labelled activity  $\alpha^{P \rightarrow Q}$  in state  $\mathbf{x}$  is defined as

$$f(\mathbf{x}, \alpha^{P \rightarrow Q}) = \mathbf{x}[P]r_\alpha^{P \rightarrow Q}. \quad (3)$$

2. If  $\alpha$  is synchronised, with  $\text{pre}(\alpha) = \{P_1, P_2, \dots, P_k\}$ , then for each

$$(Q_1, Q_2, \dots, Q_k) \in \text{post}(P_1, \alpha) \times \text{post}(P_2, \alpha) \times \dots \times \text{post}(P_k, \alpha),$$

let  $w = (P_1 \rightarrow Q_1, P_2 \rightarrow Q_2, \dots, P_k \rightarrow Q_k)$ . Then the transition rate function of labelled activity  $\alpha^w$  in state  $\mathbf{x}$  is defined as

$$f(\mathbf{x}, \alpha^w) = \left( \prod_{i=1}^k \frac{r_{\alpha}^{P_i \rightarrow Q_i}}{r_{\alpha}(P_i)} \right) \min_{i \in \{1, \dots, k\}} \{r_{\alpha}(\mathbf{x}, P_i)\},$$

where  $r_{\alpha}(\mathbf{x}, P_i) = \mathbf{x}[P_i]r_{\alpha}(P_i)$  is the apparent rate of  $\alpha$  in  $P_i$  in state  $\mathbf{x}$ . So

$$f(\mathbf{x}, \alpha^w) = \left( \prod_{i=1}^k \frac{r_{\alpha}^{P_i \rightarrow Q_i}}{r_{\alpha}(P_i)} \right) \min_{i \in \{1, \dots, k\}} \{\mathbf{x}[P_i]r_{\alpha}(P_i)\}. \quad (4)$$

Note that Definition 6 accommodates the passive or unspecified rate  $\top$ . An algorithm for automatically deriving the numerical representation of a PEPA model was presented in [22].

**Remark 1.** Definition 6 accommodates the passive or unspecified rate  $\top$ . If there are some  $r_l^{U \rightarrow V}$  which are  $\top$ , then the relevant calculation in the rate functions (3) and (4) can be made according to the following inequalities and equations that define the comparison and manipulation of unspecified activity rates (see Section 3.3.5 in [1]):

$$\begin{array}{ll} r < w\top & \text{for all } r \in \mathbb{R}^+ \text{ and for all } w \in \mathbb{N} \\ w_1\top < w_2\top & \text{if } w_1 < w_2 \text{ for all } w_1, w_2 \in \mathbb{N} \\ w_1\top + w_2\top = (w_1 + w_2)\top & \text{for all } w_1, w_2 \in \mathbb{N} \\ \frac{w_1\top}{w_2\top} = \frac{w_1}{w_2} & \text{for all } w_1, w_2 \in \mathbb{N} \end{array}$$

Moreover, we assume that  $0 \cdot \top = 0$ . So the terms such as “ $\min\{A\top, rB\}$ ” are interpreted as [23]:

$$\min\{A\top, rB\} = \begin{cases} rB, & A > 0, \\ 0, & A = 0. \end{cases}$$

The transition rate function has the following properties [22]:

**Proposition 1.** *The transition rate function is nonnegative; if  $P$  is a pre local derivative of  $l$ , i.e.  $P \in \text{pre}(l)$ , then the transition rate function of  $l$  in a state  $\mathbf{x}$  is less than the apparent rate of  $l$  in  $U$  in this state, that is*

$$0 \leq f(\mathbf{x}, l) \leq r_l(\mathbf{x}, P) = \mathbf{x}[P]r_l(P),$$

where  $r_l(P)$  is the apparent rate of  $l$  in  $P$  for a single instance of  $P$ .

**Proposition 2.** *Let  $l$  be an labelled activity, and  $\mathbf{x}, \mathbf{y}$  be two states. The transition rate function  $f(\mathbf{x}, l)$  defined in Definition 6 satisfies:*

1. *For any  $H > 0$ ,  $Hf(\mathbf{x}/H, l) = f(\mathbf{x}, l)$ .*
2. *There exists  $M > 0$  such that  $|f(\mathbf{x}, l) - f(\mathbf{y}, l)| \leq M\|\mathbf{x} - \mathbf{y}\|$  for any  $\mathbf{x}, \mathbf{y}$  and  $l$ .*

Hereafter  $\|\cdot\|$  denotes any matrix norm since all finite matrix norms are equivalent. The first term of this proposition illustrates a homogenous property of the rate function, while the second indicates the Lipschitz continuous property, both with respect to states.

### 3. Fluid approximations for PEPA models

The section will introduce the fluid-flow approximations for PEPA models, which leads to some kind of nonlinear ODEs. The existence and uniqueness of the solutions of the ODEs will be established. Moreover, a conservation law satisfied by the ODEs will be shown.

#### 3.1. State space explosion problem: an illustration by a tiny example

Let us first consider the following tiny example.

**Model 1.**

$$\begin{aligned}
& User_1 \stackrel{\text{def}}{=} (task_1, a).User_2 \\
& User_2 \stackrel{\text{def}}{=} (task_2, b).User_1 \\
& Provider_1 \stackrel{\text{def}}{=} (task_1, a).Provider_2 \\
& Provider_2 \stackrel{\text{def}}{=} (reset, d).Provider_1 \\
& \underbrace{User_1 || \dots || User_1}_{M \text{ copies}} \bowtie_{\{task_1\}} \underbrace{Provider_1 || \dots || Provider_1}_{N \text{ copies}}
\end{aligned}$$

According to the semantics of PEPA originally defined in [1], the size of the state space of the CTMC underlying Model 1 is  $2^{M+N}$ . That is, the size of the state space increases exponentially with the numbers of the users and providers in the system. Consequently, the dimension of the infinitesimal generator of the CTMC is  $2^{M+N} \times 2^{M+N}$ . The computational complexity of solving the global balance equation to get the steady-state probability distribution and thus derive the system performance, is therefore exponentially increasing with the numbers of the components. When  $M$  and/or  $N$  are large, the calculation of the stationary probability distribution will be infeasible due to limited resources of memory and time. The problem encountered here is the so-called *state-space explosion* problem.

A model aggregation technique, i.e. representing the states by numerical vector forms, which is introduced by Gilmore *et al.* in [24] and by Hillston in [21], can help to relieve the state-space explosion problem. It has been proved in [22] that by employing numerical vector forms the size of the state space can be reduced to  $2^M \times 2^N$  to  $(M+1) \times (N+1)$ , without relevant information and accuracy loss. However, this does not imply there is no complexity problem. The following table, Table 1, gives the runtimes of deriving the state space in several different scenarios. All experiments were carried out using the PEPA Plug-in (v0.0.19) for Eclipse Platform (v3.4.2), on a 2.66GHz Xeon CPU with 4Gb RAM running Scientific Linux 5. The runtimes here are elapsed times reported by the Eclipse platform.

Table 1: Elapsed time of state space derivation

$(M, N)$	(300,300)	(350,300)	(400,300)	(400,400)
time	2879 ms	4236 ms	“Java heap space”	“GC overhead limit exceeded”

If there are 400 users and 300 providers in the system, the Eclipse platform reports the error message of “Java heap space”, while 400 users and 400 providers result in the error information of “GC overhead limit exceeded”. These experiments show that the state-space explosion problem cannot be completely solved by just using the technique of numerical vector form, even for a tiny PEPA model. That is, in order to do practical analysis for large scale PEPA models we need new approaches.

### 3.2. Fluid approximation of PEPA models

In the numerical representation of PEPA presented in Section 2.2, a numerical vector form is introduced to capture the state information of models with repeated components. In this vector form there is one entry for each local derivative of each component type in the model. The entries in the vector are no longer syntactic terms representing the local derivative of the sequential component, but the number of components currently exhibiting this local derivative. Each numerical vector represents a single state of the system. The rates of the transitions between states are specified by the transition rate functions. For example, the transition from state  $\mathbf{x}$  to  $\mathbf{x} + l$  can be written as

$$\mathbf{x} \xrightarrow{(l, f(\mathbf{x}, l))} \mathbf{x} + l,$$

where  $l$  is a transition vector corresponding to the labelled activity  $l$  (for convenience, hereafter each pair of transition vectors and corresponding labelled activities shares the same notation), and  $f(\mathbf{x}, l)$  is the transition rate function, reflecting the intensity of the transition from  $\mathbf{x}$  to  $\mathbf{x} + l$ .

The state space is inherently discrete with the entries within the numerical vector form always being non-negative integers and always being incremented or decremented in steps of one. As pointed out in [21], when the numbers of components are large these steps are relatively small and we can approximate the behaviour by considering the movement between states to be continuous, rather than occurring in discontinuous jumps. In fact, let us consider the evolution of the numerical state vector. Denote the state at time  $t$  by  $\mathbf{x}(t)$ . In a short time  $\Delta t$ , the change to the vector  $\mathbf{x}(t)$  will be

$$\mathbf{x}(\cdot, t + \Delta t) - \mathbf{x}(\cdot, t) = F(\mathbf{x}(\cdot, t))\Delta t = \Delta t \sum_{l \in \mathcal{A}_{\text{label}}} lf(\mathbf{x}(\cdot, t), l).$$

Dividing by  $\Delta t$  and taking the limit,  $\Delta t \rightarrow 0$ , we obtain a set of ordinary differential equations (ODEs):

$$\frac{d\mathbf{x}}{dt} = F(\mathbf{x}), \tag{5}$$

where

$$F(\mathbf{x}) = \sum_{l \in \mathcal{A}_{\text{label}}} lf(\mathbf{x}, l). \tag{6}$$

Once the activity matrix and the transition rate functions are generated, the ODEs are immediately available. All of them can be obtained automatically by a derivation algorithm presented in [22].

Let  $U$  be a local derivative. For any transition vector  $l$ ,  $l[U]$  is either  $\pm 1$  or 0. If  $l[U] = -1$  then  $U$  is in the pre set of  $l$ , i.e.  $U \in \text{pre}(l)$ , while  $l[U] = 1$  implies  $U \in \text{post}(l)$ . According to (5) and (6),

$$\begin{aligned} \frac{d\mathbf{x}(U, t)}{dt} &= \sum_l l[U] f(\mathbf{x}, l) \\ &= - \sum_{l: l[U] = -1} f(\mathbf{x}, l) + \sum_{l: l[U] = 1} f(\mathbf{x}, l) \\ &= - \sum_{\{l | U \in \text{pre}(l)\}} f(\mathbf{x}, l) + \sum_{\{l | U \in \text{post}(l)\}} f(\mathbf{x}, l). \end{aligned} \tag{7}$$

The term  $\sum_{\{l | U \in \text{pre}(l)\}} f(\mathbf{x}, l)$  represents the “exit rates” in the local derivative  $U$ , while the term  $\sum_{\{l | U \in \text{post}(l)\}} f(\mathbf{x}, l)$  reflects the “entry rates” in  $U$ . The formulae (5) and (6) are activity centric while (7) is local derivative centric. Our approach to derive ODEs has extended previous results presented in the literature [21, 23, 25], by relaxing restrictions such as allowing shared activities may have different local rates, each action name may appear in different local derivatives within the definition of a sequential component, and may occur multiple times with that derivative definition, etc.

For an arbitrary CTMC, the evolution of probabilities distributed on each state can be described by a set of linear ODEs ([26], page 52). For example, for the (aggregated) CTMC underlying a PEPA model, the corresponding differential equations describing the evolution of the probability distributions are

$$\frac{d\pi}{dt} = Q^T \pi, \tag{8}$$

where each entry of  $\pi(t)$  represents the probability of the system being in each state at time  $t$ , and  $Q$  is an infinitesimal generator matrix corresponding to the CTMC. Clearly, the dimension of the coefficient matrix  $Q$  is the square of the size of the state space, which increases as the number of components increases.

The derived ODEs (5) describe the evolution of the population of the components in *each local derivative*, while (8) reflects the the probability evolution at *each state*. Since the scale of (5), i.e. the number of the ODEs, is only determined by the number of local derivatives and is unaffected by the size of the state space, so it avoids the state-space explosion problem. But the scale of (8) depends on the size of the state space, so it suffers from

the explosion problem. The price paid is that the ODEs (5) are generally nonlinear due to synchronisations, while (8) is linear. However, if there is no synchronisation contained then (5) becomes linear, and there is some correspondence and consistency between these two different types of ODEs, which will be demonstrated in Section 3.4.

### 3.3. Existence and uniqueness of ODEs' solution

For any set of ODEs, it is important to consider if the equations have a solution, and if so whether that solution is unique.

**Theorem 1.** *For a given PEPA model without passive rates, the derived ODEs from this model have a unique solution in the time interval  $[0, \infty)$ .*

*Proof.* Notice that each entry of  $F(\mathbf{x}) = \sum_l lf(\mathbf{x}, l)$  is a linear combination of the transition rate functions  $f(\mathbf{x}, l)$ , so  $F(\mathbf{x})$  is globally Lipschitz continuous since each  $f(\mathbf{x}, l)$  is globally Lipschitz continuous. That is, there exists  $M > 0$  such that  $\forall \mathbf{x}, \mathbf{y}$ ,

$$\|F(\mathbf{x}) - F(\mathbf{y})\| \leq M\|\mathbf{x} - \mathbf{y}\|. \quad (9)$$

By the classical theory in ODEs (e.g. Theorem 6.2.3 in [27], page 14), the derived ODEs have a unique solution in  $[0, \infty)$ .  $\square$

As we have mentioned, in the formula (7), the term  $\sum_{\{l|U \in \text{pre}(l)\}} f(\mathbf{x}, l)$  represents the exit rates in the local derivative  $U$ , while the term  $\sum_{\{l|U \in \text{post}(l)\}} f(\mathbf{x}, l)$  reflects the entry rates in  $U$ . For each type of component at any time, the sum of all exit activity rates must be equal to the sum of all entry activity rates, since the system is closed and there is no exchange with the environment. This leads to the following proposition.

**Proposition 3.** *Let  $C_{i_j}$  be a local derivative of component type  $C_i$ . Then for any  $i$  and  $t$ ,  $\sum_j \frac{d\mathbf{x}(C_{i_j}, t)}{dt} = 0$ , and  $\sum_j \mathbf{x}(C_{i_j}, t) = \sum_j \mathbf{x}(C_{i_j}, 0)$ .*

*Proof.* In the definition of activity matrix, the numbers of  $-1$  and  $1$  appearing in the entries of any transition vector  $l$  (i.e. a column of the activity matrix), which correspond to the component type  $C_i$ , are the same [22], i.e.

$$\#\{j : l[C_{i_j}] = -1\} = \#\{j : l[C_{i_j}] = 1\}. \quad (10)$$

Let  $\mathbf{y}$  be an indicator vector with the same dimension as  $l$  satisfying:

$$\mathbf{y}[C_{i_j}] = \begin{cases} 1, & \text{if } l[C_{i_j}] = \pm 1, \\ 0, & \text{otherwise.} \end{cases}$$

So  $\mathbf{y}^T l = 0$  by (10). Thus

$$\mathbf{y}^T \frac{d\mathbf{x}}{dt} = \mathbf{y}^T \sum_l l f(\mathbf{x}, l) = \sum_l \mathbf{y}^T l f(\mathbf{x}, l) = 0.$$

That is,  $\sum_j \frac{d\mathbf{x}(C_{i_j}, t)}{dt} = \mathbf{y}^T \frac{d\mathbf{x}}{dt} = 0$ . So  $\sum_j \mathbf{x}(C_{i_j}, t)$  is a constant and equal to  $\sum_j \mathbf{x}(C_{i_j}, 0)$ , i.e. the number of the copies of component type  $C_i$  in the system initially.  $\square$

Proposition 3 means that the ODEs satisfy a *Conservation Law*, i.e. the number of each kind of component remains constant at all times.

### 3.4. Convergence and consistence of ODEs' solution: nonsynchronised models

Now we consider PEPA models without synchronisation. For this special class of PEPA models, we will show that the solutions of the derived ODEs have finite limits. Moreover, the limits coincide with the steady-state probability distributions of the underlying CTMCs.

#### 3.4.1. Features of ODEs without synchronisations

Suppose the PEPA model has no synchronisations. Without loss of generality, we suppose that there is only one kind of component  $C$  in the system. In fact, if there are several types of component in the system, the ODEs related to the different types of component can be separated and treated independently since there are no interactions between them. Thus, we assume there is only one kind of component  $C$  and that  $C$  has  $k$  local derivatives:  $C_1, C_2, \dots, C_k$ . Then (5) is

$$\frac{d(\mathbf{x}(C_1, t), \dots, \mathbf{x}(C_k, t))^T}{dt} = \sum_l l f(\mathbf{x}, l). \quad (11)$$

Since (11) are linear ODEs, we may rewrite (11) as the following matrix form:

$$\frac{d(\mathbf{x}(C_1, t), \dots, \mathbf{x}(C_k, t))^T}{dt} = Q^T (\mathbf{x}(C_1, t), \dots, \mathbf{x}(C_k, t))^T, \quad (12)$$



where  $Q = (q_{ij})$  is a  $k \times k$  matrix.

$Q$  has many good properties.

**Proposition 4.**  $Q = (q_{ij})_{k \times k}$  in (12) is an infinitesimal generator matrix, that is,  $(q_{ij})_{k \times k}$  satisfies

1.  $0 \leq -q_{ii} < \infty$  for all  $i$ ;
2.  $q_{ij} \geq 0$  for all  $i \neq j$ ;
3.  $\sum_{j=1}^k q_{ij} = 0$  for all  $i$ .

*Proof.* According to (12), we have

$$\frac{d\mathbf{x}(C_i, t)}{dt} = \sum_{j=1}^k \mathbf{x}(C_j, t) q_{ji}. \quad (13)$$

Notice by (7),

$$\frac{d\mathbf{x}(C_i, t)}{dt} = - \sum_{\{l|C_i \in \text{pre}(l)\}} f(\mathbf{x}, l) + \sum_{\{l|C_i \in \text{post}(l)\}} f(\mathbf{x}, l).$$

So

$$\sum_{j=1}^k \mathbf{x}(C_j, t) q_{ji} = - \sum_{\{l|C_i \in \text{pre}(l)\}} f(\mathbf{x}, l) + \sum_{\{l|C_i \in \text{post}(l)\}} f(\mathbf{x}, l). \quad (14)$$

Since there is no synchronisation in the system, the transition function  $f(\mathbf{x}, l)$  is linear with respect to  $\mathbf{x}$  and therefore there is no nonlinear term, “min”, in it. In particular, if  $C_i \in \text{pre}(l)$ , then  $f(\mathbf{x}, l) = r_l(C_i) \mathbf{x}[C_i]$ , which is the apparent rate of  $l$  in  $C_i$  in state  $\mathbf{x}$  defined in Definition 5. We should point out that according to our semantics of mapping PEPA models to ODEs, the fluid approximation-version of  $f(\mathbf{x}, l)$  also holds, i.e.  $f(\mathbf{x}(t), l) = r_l(C_i) \mathbf{x}(C_i, t)$ . So (14) becomes

$$\mathbf{x}(C_i, t) q_{ii} + \sum_{j \neq i} \mathbf{x}(C_j, t) q_{ji} = \mathbf{x}(C_i, t) \sum_{\{l|C_i \in \text{pre}(l)\}} (-r_l(C_i)) + \sum_{\{l|C_i \in \text{post}(l)\}} f(\mathbf{x}, l). \quad (15)$$

Moreover, as long as  $f(\mathbf{x}, l) = r_l(C_i) \mathbf{x}(C_i)$  for some  $l$  and some positive constants  $r_l(C_i)$ , which implies that  $l$  can be fired at  $C_i$ , we must have  $C_i \in \text{pre}(l)$ . That is to say, if  $C_i \in \text{post}(l)$  then  $f(\mathbf{x}, l)$  cannot be of the form

of  $r\mathbf{x}(C_i, t)$  for any constant  $r > 0$ . Otherwise, we have  $C_i \in \text{pre}(l)$ , which results a contradiction<sup>1</sup> to  $C_i \in \text{post}(l)$ . So according to (15), we have

$$\mathbf{x}(C_i, t)q_{ii} = \mathbf{x}(C_i, t) \sum_{\{l|C_i \in \text{pre}(l)\}} (-r_l(C_i)), \quad (16)$$

$$\sum_{j \neq i} \mathbf{x}(C_j, t)q_{ji} = \sum_{\{l|C_i \in \text{post}(l)\}} f(\mathbf{x}, l). \quad (17)$$

Thus by (16),  $q_{ii} = \sum_{\{l|C_i \in \text{pre}(l)\}} (-r_l(C_i))$ , and  $0 \leq -q_{ii} < \infty$  for all  $i$ . Item 1 is proved.

Similarly, for any  $C_j$ ,  $j \neq i$ , if  $f(\mathbf{x}, l) = r\mathbf{x}(C_j, t)$  for some  $l$  and positive constant  $r$ , then clearly  $C_j$  is in the pre set of  $l$ . That is  $C_j \in \text{pre}(l)$ . So by (17),

$$\mathbf{x}(C_j, t)q_{ji} = \sum_{\{l|C_j \in \text{pre}(l), C_i \in \text{post}(l)\}} f(\mathbf{x}, l) = \mathbf{x}(C_j, t) \sum_l r_l^{C_j \rightarrow C_i}, \quad (18)$$

which implies  $q_{ji} = \sum_l r_l^{C_j \rightarrow C_i} \geq 0$  for all  $i \neq j$ , i.e. item 2 holds.

We now prove item 3. By Proposition 3,

$$\frac{\mathbf{x}(C_1, t)}{dt} + \frac{\mathbf{x}(C_2, t)}{dt} + \dots + \frac{\mathbf{x}(C_k, t)}{dt} = 0. \quad (19)$$

Then by (13) and (19), for all  $t$ ,

$$\begin{aligned} & \mathbf{x}(C_1, t) \sum_{j=1}^k q_{1j} + \mathbf{x}(C_2, t) \sum_{j=1}^k q_{2j} + \dots + \mathbf{x}(C_k, t) \sum_{j=1}^k q_{kj} \\ &= \sum_{i=1}^k \mathbf{x}(C_i, t) \sum_{j=1}^k q_{ij} = \sum_{j=1}^k \sum_{i=1}^k \mathbf{x}(C_i, t) q_{ij} = \sum_{j=1}^k \frac{d\mathbf{x}(C_j, t)}{dt} = 0. \end{aligned}$$

This implies  $\sum_{j=1}^k q_{ij} = 0$  for all  $i$ . □

---

<sup>1</sup>In this paper we do not allow a self-loop in the considered model. That is, any PEPA definition like " $C \stackrel{\text{def}}{=} (\alpha, r).C$ " which results in  $C \in \text{pre}(\alpha)$  and  $C \in \text{post}(\alpha)$  simultaneously, is not allowed.

In the proof of Proposition 4, we have shown the relationship between the coefficient matrix  $Q$  and the activity rates:

$$q_{ii} = - \sum_{\{l|C_i \in \text{pre}(l)\}} r_l(C_i), \quad q_{ij} = \sum_l r_l^{C_i \rightarrow C_j} \quad (i \neq j).$$

We point out that this infinitesimal generator matrix  $Q_{k \times k}$  may not be the infinitesimal generator matrix of the CTMC derived via the usual semantics of PEPA (we call it the “original” CTMC for convenience). In fact, the original CTMC has a state space with  $k^N$  states and the dimension of its infinitesimal generator matrix is  $k^N \times k^N$ , where  $N$  is the total number of components in the system. However, this  $Q_{k \times k}$  is the infinitesimal generator matrix of a CTMC underlying the PEPA model in which there is only one copy of the component, i.e.  $N = 1$ . To distinguish this from the original one, we refer to this CTMC as the “*singleton*” CTMC.

#### 3.4.2. Convergence and consistency for the ODEs

Proposition 4 illustrates that the coefficient matrix of the derived ODEs is an infinitesimal generator. If there is only one component in the system, then equation (12) captures the probability distribution evolution equations of the original CTMC. Based on this proposition, we can furthermore determine the convergence of the solutions.

**Theorem 2.** *Suppose  $\mathbf{x}(C_j, t)$  ( $j = 1, 2, \dots, k$ ) satisfy (11), then for any given initial values  $\mathbf{x}(C_j, 0) \geq 0$  ( $j = 1, 2, \dots, k$ ), there exist constants  $\mathbf{x}(C_j, \infty)$ , such that*

$$\lim_{t \rightarrow \infty} \mathbf{x}(C_j, t) = \mathbf{x}(C_j, \infty), \quad j = 1, 2, \dots, k. \quad (20)$$

*Proof.* By Proposition 4, the matrix  $Q$  in (12) is an infinitesimal generator matrix. Consider a “singleton” CTMC which has the state space  $S = \{C_1, C_2, \dots, C_k\}$ , the infinitesimal generator matrix  $Q$  in (12) and the initial probability distribution  $\pi(C_j, 0) = \frac{\mathbf{x}(C_j, 0)}{N}$  ( $j = 1, 2, \dots, k$ ). Then according to Markov theory ([26], page 52),  $\pi(C_j, t)$  ( $j = 1, 2, \dots, k$ ), the probability distribution of this new CTMC at time  $t$ , satisfies

$$\frac{d(\pi(C_1, t), \dots, \pi(C_k, t))}{dt} = (\pi(C_1, t), \dots, \pi(C_k, t)) Q \quad (21)$$

Since the singleton CTMC is assumed irreducible and positive-recurrent, it has a steady-state probability distribution  $\{\pi(C_j, \infty)\}_{j=1}^k$ , and

$$\lim_{t \rightarrow \infty} \pi(C_j, t) = \pi(C_j, \infty), \quad j = 1, 2, \dots, k. \quad (22)$$

Note that  $\frac{\mathbf{x}(C_j, t)}{N}$  also satisfies (21) with the initial values  $\frac{\mathbf{x}(C_j, 0)}{N}$  equal to  $\pi(C_j, 0)$ , where  $N$  is the population of the components. By the uniqueness of the solutions of (21), we have

$$\frac{\mathbf{x}(C_j, t)}{N} = \pi(C_j, t), \quad j = 1, 2, \dots, k, \quad (23)$$

and hence by (22),

$$\lim_{t \rightarrow \infty} \mathbf{x}(C_j, t) = \lim_{t \rightarrow \infty} N\pi(C_j, t) = N\pi(C_j, \infty), \quad j = 1, 2, \dots, k.$$

□

Clearly, if there are multiple types of component in the system, then Theorem 2 holds for each each component type, since there is no cooperation between different component types and they can be treated independently.

It is shown in [28] that for some special examples the equilibrium solutions of the ODEs coincide with the steady state probability distributions of the underlying original CTMC. This theorem states that this holds for all for PEPA models without synchronisations. Moreover, it is also exposed by this theorem that the fluid approximation is consistent with the CTMC underlying the same nonsynchronised PEPA model.

#### 4. Relating to density dependent CTMCs

A general PEPA model may have synchronisations, which result in the nonlinearity of the derived ODEs. However, it is difficult to rely on pure analytical methods to explore the asymptotic behaviour of the solution of the derived ODEs from an arbitrary PEPA model (except for some special classes of models, see the next two sections).

Fortunately, Kurtz's theorem [29, 30] establishes the relationship between a sequence of Markov chains and a corresponding set of ODEs: the complete solution of some ODEs is the limit of a sequence of Markov chains. In the context of PEPA, the derived ODEs can be considered as the limit of pure

jump Markov processes, as first exposed in [25] for a special case. Thus we may investigate the convergence of the ODEs' solutions by alternatively studying the corresponding property of the Markov chains through this consistency relationship. This approach leads to the result presented in the next section: under a particular condition the solution will converge and the limit is consistent with the limit steady-state probability distribution of a family of CTMCs underlying the given PEPA model. Let us first introduce the concept of density dependent Markov chains underlying PEPA models.

#### 4.1. Density dependent Markov chains underlying PEPA models

In the numerical state vector representation scheme, each vector is a single state and the rates of the transitions between states are specified by the rate functions. For example, the transition from state  $\mathbf{x}$  to  $\mathbf{x} + l$  can be written as

$$\mathbf{x} \xrightarrow{(l, f(\mathbf{x}, l))} \mathbf{x} + l.$$

Since all the transitions are only determined by the current state rather than the previous ones, given any starting state a CTMC can be obtained. More specifically, the state space of the CTMC is the set of all reachable numerical state vectors  $\mathbf{x}$ . The infinitesimal generator is determined by the transition rate function,

$$q_{\mathbf{x}, \mathbf{x}+l} = f(\mathbf{x}, l). \quad (24)$$

Because the transition rate function is defined according to the semantics of PEPA, the CTMC mentioned above is in fact the aggregated CTMC underlying the given PEPA model. In other words, the transition rate of the aggregated CTMC is specified by the transition rate function in Definition 6.

It is obvious that the aggregated CTMC depends on the starting state of the given PEPA model. By altering the population of components presented in the model, which can be done by varying the initial states, we may get a sequence of aggregated CTMCs. Moreover, Proposition 2 indicates that the transition rate function has the homogenous property:  $Hf(\mathbf{x}/H, l) = f(\mathbf{x}, l)$ ,  $\forall H > 0$ . This property identifies the aggregated CTMC to be *density dependent*.

**Definition 7.** [29]. A family of CTMCs  $\{X_n\}_n$  is called density dependent if and only if there exists a continuous function  $f(\mathbf{x}, l)$ ,  $\mathbf{x} \in \mathbb{R}^d$ ,  $l \in \mathbb{Z}^d$ , such that the infinitesimal generators of  $X_n$  are given by:

$$q_{\mathbf{x}, \mathbf{x}+l}^{(n)} = nf(\mathbf{x}/n, l), \quad l \neq 0,$$

where  $q_{\mathbf{x}, \mathbf{x}+l}^{(n)}$  denotes an entry of the infinitesimal generator of  $X_n$ ,  $\mathbf{x}$  a numerical state vector and  $l$  a transition vector.

This allows us to conclude the following proposition.

**Proposition 5.** *Let  $\{X_n\}$  be a sequence of aggregated CTMCs generated from a given PEPA model (by scaling the initial state), then  $\{X_n\}$  is density dependent.*

*Proof.* For any  $n$ , the transition between states is determined by

$$q_{\mathbf{x}, \mathbf{x}+l}^{(n)} = f(\mathbf{x}, l),$$

where  $\mathbf{x}, \mathbf{x} + l$  are state vectors,  $l$  corresponds to an activity,  $f(\mathbf{x}, l)$  is the rate of the transition from state  $\mathbf{x}$  to  $\mathbf{x} + l$ . By Proposition 2,

$$nf(\mathbf{x}/n, l) = f(\mathbf{x}, l).$$

So the infinitesimal generator of  $X_n$  is given by:

$$q_{\mathbf{x}, \mathbf{x}+l}^{(n)} = f(\mathbf{x}, l) = nf(\mathbf{x}/n, l), \quad l \neq 0.$$

Therefore,  $\{X_n\}$  is a sequence of density dependent CTMCs.  $\square$

In particular, the family of density dependent CTMCs,  $\{X_n(t)\}$ , derived from a given PEPA model with the starting condition  $X_n(0) = n\mathbf{x}_0$  ( $\forall n$ ), is called the *density dependent CTMCs associated with  $\mathbf{x}_0$* . The CTMCs  $\frac{X_n(t)}{n}$  are called the *concentrated density dependent CTMCs*. Here  $n$  is called the *concentration level*, indicating that the entries within the numerical vector states (of  $\frac{X_n(t)}{n}$ ) are incremented and decremented in steps of  $\frac{1}{n}$ .

For example, Consider the following PEPA model, which is Model 1 presented previously:

$$\begin{aligned} & User_1 \stackrel{def}{=} (task_1, a).User_2 \\ & User_2 \stackrel{def}{=} (task_2, b).User_1 \\ & Provider_1 \stackrel{def}{=} (task_1, a).Provider_2 \\ & Provider_2 \stackrel{def}{=} (reset, d).Provider_1 \\ & (User_1[M]) \bowtie_{\{task_1\}} (Provider_1[N]). \end{aligned}$$

The activity matrix and transition rate functions have been specified in Table 2. In this table,  $U_i, P_i$  ( $i = 1, 2$ ) are the local derivatives representing  $User_i$  and  $Provider_i$  respectively. For convenience, the labelled activities or transition vectors  $task_1^{(U_1 \rightarrow U_2, P_1 \rightarrow P_2)}$ ,  $task_2^{U_2 \rightarrow U_1}$ ,  $reset^{P_2 \rightarrow P_1}$  will subsequently be denoted by  $l^{task_1}$ ,  $l^{task_2}$ ,  $l^{reset}$  respectively.

Table 2: Activity matrix and transition rate function of Model 1

$l$	$task_1^{(U_1 \rightarrow U_2, P_1 \rightarrow P_2)}$	$task_2^{U_2 \rightarrow U_1}$	$reset^{P_2 \rightarrow P_1}$
$U_1$	-1	1	0
$U_2$	1	-1	0
$P_1$	-1	0	1
$P_2$	1	0	-1
$f(\mathbf{x}, l)$	$a \min(\mathbf{x}[U_1], \mathbf{x}[P_1])$	$b\mathbf{x}[U_2]$	$d\mathbf{x}[P_2]$

Suppose  $\mathbf{x}_1 = (M, 0, N, 0)^T = (1, 0, 1, 0)^T$ . Let  $X_1(t)$  be the aggregated CTMC underlying Model 1 with initial state  $\mathbf{x}_1$ . Then the state space of  $X_1(t)$ , denoted by  $S_1$ , is composed of

$$\begin{aligned} \mathbf{x}_1 &= (1, 0, 1, 0)^T, & \mathbf{x}_2 &= (0, 1, 0, 1)^T, \\ \mathbf{x}_3 &= (1, 0, 0, 1)^T, & \mathbf{x}_4 &= (0, 1, 1, 0)^T. \end{aligned} \quad (25)$$

According to the transition rate functions presented in Table 2, we have, for instance,

$$q_{\mathbf{x}_1, \mathbf{x}_2}^{(1)} = q_{\mathbf{x}_1, \mathbf{x}_1 + l^{task_1}} = f(\mathbf{x}_1, l^{task_1}) = a \min(\mathbf{x}_1[U_1], \mathbf{x}_1[P_1]) = a.$$

Varying the initial states we may get other aggregated CTMCs. For example, let  $X_2(t)$  be the aggregated CTMC corresponding to the initial state  $X_2(0) = 2\mathbf{x}_0 = (2, 0, 2, 0)^T$ . Then the state space  $S_2$  of  $X_2(t)$  has the states

$$\begin{aligned} \mathbf{x}_1 &= (2, 0, 2, 0)^T, & \mathbf{x}_2 &= (1, 1, 1, 1)^T, & \mathbf{x}_3 &= (1, 1, 2, 0)^T, \\ \mathbf{x}_4 &= (1, 1, 0, 2)^T, & \mathbf{x}_5 &= (0, 2, 1, 1)^T, & \mathbf{x}_6 &= (2, 0, 1, 1)^T, \\ \mathbf{x}_7 &= (0, 2, 0, 2)^T, & \mathbf{x}_8 &= (0, 2, 2, 0)^T, & \mathbf{x}_9 &= (2, 0, 0, 2)^T. \end{aligned} \quad (26)$$

The rate of transition from  $\mathbf{x}_1$  to  $\mathbf{x}_2$  is determined by

$$q_{\mathbf{x}_1, \mathbf{x}_2}^{(2)} = q_{\mathbf{x}_1, \mathbf{x}_1 + l^{task_1}} = f(\mathbf{x}_1, l^{task_1}) = 2a = 2f(\mathbf{x}_1/2, l^{task_1}).$$

Similarly, let  $X_n(t)$  be the aggregated CTMC corresponding to the initial state  $X_n(0) = n\mathbf{x}_0$ . Then the transition from  $\mathbf{x}$  to  $\mathbf{x} + l$  is determined by

$$q_{\mathbf{x}, \mathbf{x}+l}^{(n)} = f(\mathbf{x}, l) = nf(\mathbf{x}/n, l).$$

Thus a family of aggregated CTMCs, i.e.  $\{X_n(t)\}$ , has been obtained from Model 1. These derived  $\{X_n(t)\}$  are density dependent CTMCs associated with  $\mathbf{x}_0$ . As illustrated by this example, the density dependent CTMCs are obtained by scaling the starting state  $\mathbf{x}_0$ . So the starting state of each CTMC is different, because  $X_n(0) = n\mathbf{x}_0$ , i.e.  $X_n(0) = n(M, 0, N, 0)^T$ .

#### 4.2. Fluid approximation as the limit of the CTMCs

As discussed above, a set of ODEs and a sequence of density dependent Markov chains can be derived from the same PEPA model. The former one is deterministic while the latter is stochastic. However, both of them are determined by the same activity matrix and the same rate functions that are uniquely generated from the given PEPA model. Therefore, it is natural to believe that there is some kind of consistency between them.

As we have mentioned, the complete solution of some ODEs can be the limit of a sequence of Markov chains according to Kurtz's theorem [29, 30]. Such consistency in the context of PEPA has been previously illustrated for a particular PEPA model [25]. Here we give a modified version of this result for general PEPA models, in which the convergence is in the sense of almost surely rather than probabilistically as in [25].

**Theorem 3.** *Let  $X(t)$  be the solution of the ODEs (5) derived from a given PEPA model with initial condition  $\mathbf{x}_0$ , and let  $\{X_n(t)\}$  be the density dependent CTMCs associated with  $\mathbf{x}_0$  underlying the same PEPA model. Let  $\hat{X}_n(t) = \frac{X_n(t)}{n}$ , then for any  $t > 0$ ,*

$$\lim_{n \rightarrow \infty} \sup_{u \leq t} \|\hat{X}_n(u) - X(u)\| = 0 \quad a.s. \quad (27)$$

*Proof.* According to Kurtz's theorem, which is listed in Appendix A, it is sufficient to prove: for any compact set  $K \subset \mathbb{R}^{N_d}$ ,

1.  $\exists M_K > 0$  such that  $\|F(\mathbf{x}) - F(\mathbf{y})\| \leq M_K \|\mathbf{x} - \mathbf{y}\|$ ;
2.  $\sum_l \|l\| \sup_{\mathbf{x} \in K} f(\mathbf{x}, l) < \infty$ .



Clearly, above term 1 is satisfied. Since  $f(\mathbf{x}, l)$  is continuous by Proposition 2, it is bounded on any compact  $K$ . Notice that any entry of  $l$  takes values in  $\{1, -1, 0\}$ , so  $\|l\|$  is bounded. Thus term 2 is satisfied, which completes the proof.  $\square$

Theorem 3 allows us to investigate the properties of  $X(t)$  through studying the characteristics of the family of CTMCs  $\hat{X}_n(t) = \frac{X_n(t)}{n}$ . Notice that  $X_n(t)$  takes values in the state space which corresponds to the starting state  $n\mathbf{x}_0$ . Clearly, each state of  $\hat{X}_n(t)$  is bounded and nonnegative (more precisely, each entry in any numerical state vector is nonnegative, and bounded by  $\|\mathbf{x}_0\|$  according to the conservative law). So the ODEs' solution  $X(t)$  inherits these characteristics since  $X(t)$  is the limit of  $\hat{X}_n(t)$  as  $n$  goes to infinity. That is,  $X(t)$  is bounded and nonnegative. The proof is trivial and omitted here. Instead, a purely analytic proof of these properties will be given in Section 6.1.

#### 4.3. Consistency between the fluid approximation and the CTMCs

As shown in Theorem 3, for a given PEPA model with synchronisations, the derived ODEs can be taken as the underlying density dependent CTMC with the concentration level infinity. If the given model has no synchronisations, then by Proposition 4 and the proof of Theorem 2, the ODEs coincide the probability distribution evolution equations of the CTMC with the concentration level one, except for a scaling factor. These two conclusions embody the consistency between the fluid approximation and the CTMCs for PEPA models. Moreover, as we will see in the next section, the limit of the ODEs' solution as time tends to infinity (if it exists) is consistent with the limit of the expectations of the corresponding CTMCs.

However, it is natural to ask such a question: for a PEPA model with synchronisations, since the ODEs corresponds the CTMC with the concentration level infinity and our performance evaluation is based on the usual underlying CTMC, i.e. the CTMC with the concentration level one, what is the loss brought by using the fluid approximation approach? Although in the sense of concentration level there is a gap between one and infinity, but in practice for large scale synchronised models (i.e. models with a large number of repetitive components), the relative errors of performance measure such as throughput and average response time between the concentration level one and infinity are usually small enough to be ignored (i.e. less than 5%) [22]. Therefore, it is safe to employ the fluid approximation for large scale models.

This paper more likely concentrates on the long-time behaviour the ODEs' solution. In the following sections, we focus on the problem of whether the ODEs' solution converges as time goes to problem will be presented in the next section, which is based on the consistency relationship between the derived ODEs and the CTMCs revealed in Theorem 3.

## 5. Convergence of ODEs' solution: a probabilistic approach

Analogous to the steady-state probability distributions of the Markov chains underlying PEPA models, upon which performance measures such as throughput and utilisation can be derived, we expect the solution of the generated ODEs to have similar equilibrium conditions. In particular, if the solution has a limit as time goes to infinity we will be able to similarly obtain the performance from the steady state, i.e. the limit. Therefore, whether the solution of the derived ODEs converges becomes an important problem.

We should point out that Kurtz's theorem cannot directly apply to the problem of whether or not the solution the derived ODEs converges. This is because Kurtz's theorem only deals with the approximation between the ODEs and Markov chains during any finite time, rather than considering the asymptotic behaviour of the ODEs as time goes to infinity. This section will present our investigation and results about this problem.

### 5.1. Convergence under a particular condition

We follow the assumptions in Theorem 3. Denote the expectation of  $\hat{X}_n(t)$  as  $\hat{M}_n(t)$ , i.e.  $\hat{M}_n(t) = E[\hat{X}_n(t)]$ . For any  $t$ , the stochastic processes  $\{\hat{X}_n(t)\}_n$  converge to the deterministic  $X(t)$  when  $n$  tends to infinity, as Theorem 3 shows. It is not surprising to see that  $\{\hat{M}_n(t)\}_n$ , the expectations of  $\{\hat{X}_n(t)\}_n$ , also converge to  $X(t)$  as  $n \rightarrow \infty$ :

**Lemma 1.** *For any  $t$ ,*

$$\lim_{n \rightarrow \infty} \hat{M}_n(t) = X(t).$$

*Proof.* Since  $X(t)$  is deterministic, then  $E[X(t)] = X(t)$ . By Theorem 3, for all  $t$ ,  $\hat{X}_n(t)$  converges to  $X(t)$  almost surely as  $n$  goes to infinity. Notice that  $\hat{X}_n(t)$  is bounded (see the discussion in the previous section), then by Lebesgue's dominant convergence theorem, we have

$$\lim_{n \rightarrow \infty} E\|\hat{X}_n(t) - X(t)\| = 0.$$

Since a norm  $\|\cdot\|$  can be considered as a convex function, by Jensen's inequality (Theorem 2.2 in [31]), we have  $\|(E[\cdot])\| \leq E[\|\cdot\|]$ . Therefore,

$$\begin{aligned} \lim_{n \rightarrow \infty} \|\hat{M}_n(t) - X(t)\| &= \lim_{n \rightarrow \infty} \|E[\hat{X}_n(t)] - E[X(t)]\| \\ &\leq \lim_{n \rightarrow \infty} E\|\hat{X}_n(t) - X(t)\| \\ &= 0. \end{aligned}$$

□

Lemma 1 states that the ODEs' solution  $X(t)$  is just the limit function of the sequence of the expectation functions of the corresponding density dependent Markov chains. This provides some clues: the characteristics of the limit  $X(t)$  depend on the properties of  $\{\hat{M}_n(t)\}_n$ . Therefore, we expect to be able to investigate  $X(t)$  by studying  $\{\hat{M}_n(t)\}_n$ .

Since  $\hat{M}_n(t)$  is the expectation of the Markov chain  $\hat{X}_n(t)$ ,  $\hat{M}_n(t)$  can be expressed by a formula in which the transient probability distribution is involved. That is,

$$\hat{M}_n(t) = E[\hat{X}_n(t)] = \sum_{\mathbf{x} \in \hat{S}_n} \mathbf{x} \hat{\pi}_t^n(\mathbf{x}),$$

where  $\hat{S}_n$  is the state space,  $\hat{\pi}_t^n(\cdot)$  is the probability distribution of  $\hat{X}_n$  at time  $t$ . Let  $S_n$  and  $\pi_t^n(\cdot)$  be the state space and the probability distribution of  $X_n(t)$  respectively<sup>2</sup>. Then

$$\hat{M}_n(t) = E[\hat{X}_n(t)] = E\left[\frac{X_n(t)}{n}\right] = \sum_{\mathbf{x} \in S_n} \frac{\mathbf{x}}{n} \pi_t^n(\mathbf{x}).$$

We have assumed the Markov chains underlying PEPA models to be irreducible and positive-recurrent. Then the transient probability distributions of these Markov chains will converge to the corresponding steady-state probability distributions. We denote the steady-state probability distributions of  $X_n(t)$  and  $\hat{X}_n(t)$  as  $\pi_\infty^n(\cdot)$  and  $\hat{\pi}_\infty^n(\cdot)$  respectively. Then, we have a lemma.

**Lemma 2.** *For any  $n$ , there exists a  $\hat{M}_n(\infty)$ , such that*

$$\lim_{t \rightarrow \infty} \hat{M}_n(t) = \hat{M}_n(\infty).$$

---

<sup>2</sup>We should point out that the probability distributions of  $X_n(t)$  and  $\hat{X}_n(t)$  are the same, i.e.  $\pi_t^n(\mathbf{x}) = \hat{\pi}_t^n(\mathbf{x}/n)$ .

*Proof.*

$$\begin{aligned} & \lim_{t \rightarrow \infty} \hat{M}_n(t) \\ &= \lim_{t \rightarrow \infty} \sum_{\mathbf{x} \in S_n} \frac{\mathbf{x}}{n} \pi_t^n(\mathbf{x}) = \sum_{\mathbf{x} \in S_n} \lim_{t \rightarrow \infty} \frac{\mathbf{x}}{n} \pi_t^n(\mathbf{x}) = \sum_{\mathbf{x} \in S_n} \frac{\mathbf{x}}{n} \pi_\infty^n(\mathbf{x}) \equiv \hat{M}_n(\infty). \end{aligned}$$

□

Clearly, we also have  $\hat{M}_n(\infty) = \sum_{\mathbf{x} \in \hat{S}_n} \mathbf{x} \hat{\pi}_\infty^n(\mathbf{x})$ .

**Remark 2.** *Currently, we do not know whether the sequence  $\{\hat{M}_n(\infty)\}_n$  converges as  $n \rightarrow \infty$ . But since  $\{\hat{M}_n(\infty)\}_n$  is bounded which is due to the conservation law that PEPA models satisfy, there exists  $\{n'\} \subset \{n\}$  such that  $\{\hat{M}_{n'}(\infty)\}$  converges to a limit, namely  $\hat{M}_\infty(\infty)$ . That is*

$$\lim_{n' \rightarrow \infty} \hat{M}_{n'}(\infty) = \hat{M}_\infty(\infty).$$

Thus,

$$\lim_{n' \rightarrow \infty} \lim_{t \rightarrow \infty} \hat{M}_{n'}(t) = \lim_{n' \rightarrow \infty} \hat{M}_{n'}(\infty) = \hat{M}_\infty(\infty). \quad (28)$$

At the moment, there are two questions:

1. Whether  $\lim_{t \rightarrow \infty} \lim_{n' \rightarrow \infty} \hat{M}_{n'}(t)$  exists?
2. If  $\lim_{t \rightarrow \infty} \lim_{n' \rightarrow \infty} \hat{M}_{n'}(t)$  exists, whether

$$\lim_{t \rightarrow \infty} \lim_{n' \rightarrow \infty} \hat{M}_{n'}(t) = \lim_{n' \rightarrow \infty} \lim_{t \rightarrow \infty} \hat{M}_{n'}(t)?$$

If the answer to the first question is yes, then the solution of the ODEs converges, since by Lemma 2,  $\lim_{t \rightarrow \infty} X(t) = \lim_{t \rightarrow \infty} \lim_{n' \rightarrow \infty} \hat{M}_{n'}(t)$ . If the answer to the second question is yes, then the limit of  $X(t)$  is consistent with the stationary distributions of the Markov chains since

$$\lim_{t \rightarrow \infty} X(t) = \lim_{t \rightarrow \infty} \lim_{n' \rightarrow \infty} \hat{M}_{n'}(t) = \lim_{n' \rightarrow \infty} \lim_{t \rightarrow \infty} \hat{M}_{n'}(t) = \hat{M}_\infty(\infty).$$

In short, the positive answers to these two questions determine the convergence and consistency for the ODEs's solution, see Figure 1. Fortunately, the two answers can be guaranteed by the condition (29) in the following Proposition 6.

Figure 1: Convergence and consistency diagram for derived ODEs

$$\begin{array}{ccc}
\hat{M}_{n'}(t) & \xrightarrow[\text{Lemma 1}]{n' \rightarrow \infty} & X(t)(= \hat{M}_\infty(t)) \\
t \rightarrow \infty \downarrow \text{Lemma 2} & & \text{???} \downarrow t \rightarrow \infty \\
\hat{M}_{n'}(\infty) & \xrightarrow[\text{Remark 2}]{n' \rightarrow \infty} & \hat{M}_\infty(\infty)
\end{array}$$

**Proposition 6. (A particular condition)** *If there exist  $A, B > 0$ , such that*

$$\sup_{n'} \left\| \hat{M}_{n'}(t) - \hat{M}_{n'}(\infty) \right\| < B e^{-At}, \quad (29)$$

*then  $\lim_{t \rightarrow \infty} X(t) = \hat{M}_\infty(\infty)$ .*

*Proof.*

$$\begin{aligned}
\left\| X(t) - \hat{M}_\infty(\infty) \right\| &= \left\| \lim_{n' \rightarrow \infty} [\hat{M}_{n'}(t) - \hat{M}_{n'}(\infty)] \right\| \\
&\leq \limsup_{n' \rightarrow \infty} \left\| \hat{M}_{n'}(t) - \hat{M}_{n'}(\infty) \right\| \\
&\leq \limsup_{n' \rightarrow \infty} \left[ \sup_{n'} \left\| \hat{M}_{n'}(t) - \hat{M}_{n'}(\infty) \right\| \right] \\
&\leq \limsup_{n' \rightarrow \infty} B e^{-At} \\
&= B e^{-At} \longrightarrow 0, \text{ as } t \longrightarrow \infty.
\end{aligned}$$

So  $\lim_{t \rightarrow \infty} X(t) = \hat{M}_\infty(\infty)$ . □

Notice that  $\hat{M}_n(t) = \sum_{\mathbf{x} \in S^n} \frac{\mathbf{x}}{n} \pi_t^n(\mathbf{x})$  and  $\hat{M}_n(\infty) = \sum_{\mathbf{x} \in S^n} \frac{\mathbf{x}}{n} \pi_\infty^n(\mathbf{x})$ , so in order to estimate  $\left\| \hat{M}_n(t) - \hat{M}_n(\infty) \right\|$  in (29), we need first to estimate the difference between  $\pi_t^n$  and  $\pi_\infty^n$ .

**Lemma 3.** *If there exists  $A > 0$  and  $B_1 > 0$ , such that for any  $n'$  and all  $\mathbf{x} \in S^{n'}$ ,*

$$|\pi_t^{n'}(\mathbf{x}) - \pi_\infty^{n'}(\mathbf{x})| \leq \pi_\infty^{n'}(\mathbf{x}) B_1 e^{-At}, \quad (30)$$

*then there exists  $B > 0$  such that  $\sup_{n'} \left\| \hat{M}_{n'}(t) - \hat{M}_{n'}(\infty) \right\| < B e^{-At}$  holds.*

*Proof.* We know that  $\hat{X}_n(0) = \frac{X_n(0)}{n} = \mathbf{x}_0$  for any  $n$ . By the conservation law, the population of each entity in any state is determined by the starting state. So for any  $n'$  and all  $\mathbf{x} \in S^{n'}$ ,  $\|\mathbf{x}/n'\| \leq C_1 \sum_{P \in \mathcal{D}} \mathbf{x}_0[P] < \infty$ , where  $\mathcal{D}$  is the set of all local derivatives and  $C_1$  is a constant independent of  $n'$ . Let  $C = \sup_{n'} \max_{\mathbf{x} \in S^{n'}} \|\mathbf{x}/n'\|$ , then  $C < \infty$ .

$$\begin{aligned}
\left\| \hat{M}_{n'}(t) - \hat{M}_{n'}(\infty) \right\| &= \left\| \sum_{\mathbf{x} \in S^{n'}} \frac{\mathbf{x}}{n'} \pi_t^{n'}(\mathbf{x}) - \sum_{\mathbf{x} \in S^{n'}} \frac{\mathbf{x}}{n'} \pi_\infty^{n'}(\mathbf{x}) \right\| \\
&= \left\| \sum_{\mathbf{x} \in S^{n'}} \frac{\mathbf{x}}{n'} (\pi_t^{n'}(\mathbf{x}) - \pi_\infty^{n'}(\mathbf{x})) \right\| \\
&\leq \sup_{n'} \max_{\mathbf{x} \in S^{n'}} \left\| \frac{\mathbf{x}}{n'} \right\| \sum_{\mathbf{x} \in S^{n'}} |\pi_t^{n'}(\mathbf{x}) - \pi_\infty^{n'}(\mathbf{x})| \\
&= C \sum_{\mathbf{x} \in S^{n'}} |\pi_t^{n'}(\mathbf{x}) - \pi_\infty^{n'}(\mathbf{x})| \\
&\leq C \sum_{\mathbf{x} \in S^{n'}} \pi_\infty^{n'}(\mathbf{x}) B_1 e^{-tA} \\
&= C B_1 e^{-tA}.
\end{aligned}$$

Let  $B = C B_1$ . Then  $\sup_{n'} \left\| \hat{M}_{n'}(t) - \hat{M}_{n'}(\infty) \right\| < B e^{-At}$ .  $\square$

## 5.2. Investigation of the particular condition

This section will present the study of the particular condition (29). We will expose that the condition is related to well-known constants of Markov chains such as the spectral gap and the Log-Sobolev constant. The methods and results developed in the field of functional analysis of Markov chains are utilised to investigate the condition.

### 5.2.1. Important estimation for Markov kernel

We first give an estimation for the Markov kernel which is defined below. Let  $Q$  be the infinitesimal generator of a Markov chain  $X$  on a finite state  $S$ . Let

$$K_{ij} = \begin{cases} \frac{Q_{ij}}{m}, & i \neq j \\ 1 + \frac{Q_{ii}}{m}, & i = j \end{cases} \quad \text{where } m = \sup_i (-Q_{ii}).$$

$K$  is a transition probability matrix, satisfying

$$K(\mathbf{x}, \mathbf{y}) \geq 0, \quad \sum_{\mathbf{y}} K(\mathbf{x}, \mathbf{y}) = 1.$$

$K$  is called an Markov *kernel* ( $K$  is also called the *uniformisation* of the CTMC in some literature). A Markov chain on a finite state space  $S$  can be described through its *kernel*  $K$ . The continuous time semigroup associated with  $K$  is defined by

$$H_t = \exp(-t(I - K)).$$

Let  $\pi$  be the unique stationary measure of the Markov chain. Then  $H_t(\mathbf{x}, \mathbf{y}) \rightarrow \pi(\mathbf{y})$  as  $t$  tends to infinity. Following the convention in the literature we will also use  $(K, \pi)$  to represent a Markov chain.

Notice

$$Q = m(K - I), \quad K = \frac{Q}{m} + I.$$

Clearly,

$$P_t = e^{tQ} = e^{mt(K-I)} = e^{-mt(I-K)} = H_{mt},$$

and thus  $H_t = P_{\frac{t}{m}}$ , where  $P_t$  is called the semigroup associated with the infinitesimal generator  $Q$ . An estimation of  $H_t$  is given below.

**Lemma 4.** (Corollary 2.2.6, [32]) *Let  $(K, \pi)$  be a finite Markov chain, and  $\pi(*) = \min_{\mathbf{x} \in S} \pi(\mathbf{x})$ . Then*

$$\sup_{\mathbf{x}, \mathbf{y}} \left| \frac{H_t(\mathbf{x}, \mathbf{y})}{\pi(\mathbf{y})} - 1 \right| \leq e^{2-c} \quad \text{for } t = \frac{1}{\alpha} \log \log \frac{1}{\pi(*)} + \frac{c}{\lambda}, \quad (31)$$

where  $\lambda > 0, \alpha > 0$  are the spectral gap and the Log-Sobolev constant respectively, which are defined and interpreted in Appendix B.

It can be implied from (31) that  $\forall \mathbf{x}, \mathbf{y} \in S$ ,

$$|H_t(\mathbf{x}, \mathbf{y}) - \pi(\mathbf{y})| \leq \pi(\mathbf{y}) \left[ e^{2 + \frac{\lambda}{\alpha} \log \log \frac{1}{\pi(*)}} \right] e^{-\lambda t}. \quad (32)$$

Since  $H_t = P_{\frac{t}{m}}$ , so

$$\left| P_{\frac{t}{m}}(\mathbf{x}, \mathbf{y}) - \pi(\mathbf{y}) \right| \leq \pi(\mathbf{y}) \left[ e^{2 + \frac{\lambda}{\alpha} \log \log \frac{1}{\pi(*)}} \right] e^{-\lambda t}, \quad (33)$$

and thus (replacing “ $t$ ” by “ $mt$ ” on the both sides of (33)),

$$|P_t(\mathbf{x}, \mathbf{y}) - \pi(\mathbf{y})| \leq \pi(\mathbf{y}) \left[ e^{2 + \frac{\lambda}{\alpha} \log \log \frac{1}{\pi(*)}} \right] e^{-m\lambda t}. \quad (34)$$

The investigation and utilisation of the formulae (34) will be presented in the next subsection.

### 5.2.2. Study of the particular condition

For each  $n$ , let  $Q^n$  be the infinitesimal generator of the density dependent Markov chain  $X_n(t)$  underlying a given PEPA model and thus the transition probability matrix is  $P_t^n = e^{tQ^n}$ . For each  $X_n(t)$ , the initial state of the corresponding system is  $n\mathbf{x}_0$ , so the initial probability distribution of  $X_n(t)$  is  $\pi_0^n(n\mathbf{x}_0) = 1$  and  $\pi_0^n(\mathbf{x}) = 0$ ,  $\forall \mathbf{x} \in S^n, \mathbf{x} \neq n\mathbf{x}_0$ . So  $\forall \mathbf{y} \in S^n$ ,

$$\pi_t^n(\mathbf{y}) = \sum_{\mathbf{x} \in S^n} \pi_0^n(\mathbf{x}) P_t^n(\mathbf{x}, \mathbf{y}) = \pi_0^n(n\mathbf{x}_0) P_t^n(n\mathbf{x}_0, \mathbf{y}) = P_t^n(n\mathbf{x}_0, \mathbf{y}). \quad (35)$$

The formula (34) in the context of  $X_n(t)$  is

$$|P_t^n(\mathbf{x}, \mathbf{y}) - \pi_\infty^n(\mathbf{y})| \leq \pi_\infty^n(\mathbf{y}) \left[ e^{2 + \frac{\lambda_n}{\alpha_n} \log \log \frac{1}{\pi_\infty^n(*)}} \right] e^{-m_n \lambda_n t}, \quad (36)$$

where  $\lambda_n, \lambda_n, m_n, \pi_\infty^n(*)$  are the respective parameters associated with  $X_n(t)$ .

Let  $\mathbf{x} = n\mathbf{x}_0$  in (36), and notice  $P_t^n(n\mathbf{x}_0, \mathbf{y}) = \pi_t^n(\mathbf{y})$  by (35), then we have

$$|\pi_t^n(\mathbf{y}) - \pi_\infty^n(\mathbf{y})| \leq \pi_\infty^n(\mathbf{y}) \left[ e^{2 + \frac{\lambda_n}{\alpha_n} \log \log \frac{1}{\pi_\infty^n(*)}} \right] e^{-m_n \lambda_n t}, \quad \forall \mathbf{y} \in S^n. \quad (37)$$

From the comparison of (37) and (30), we know that if there are some conditions imposed on  $\frac{\lambda_n}{\alpha_n} \log \log \frac{1}{\pi_\infty^n(*)}$  and  $-m_n \lambda_n t$ , then (30) can be induced from (37). See the following Lemma.

**Lemma 5.** *If there exists  $T > 0, B_2 > 0, A > 0$  such that*

$$\sup_n \left\{ -m_n \lambda_n T + \frac{\lambda_n}{\alpha_n} \log \log \frac{1}{\pi_\infty^n(*)} \right\} \leq B_2, \quad (38)$$

and

$$\inf_n \{m_n \lambda_n\} \geq A, \quad (39)$$

then

$$|\pi_t^n(\mathbf{x}) - \pi_\infty^n(\mathbf{x})| \leq \pi_\infty^n(\mathbf{x}) B_1 e^{-At}, \quad \forall \mathbf{x} \in S^n, \quad (40)$$

where  $B_1 = e^{A_1 T + B_2 + 2}$ , and the “particular condition” (29) holds.

*Proof.*

$$\begin{aligned} \left[ e^{2 + \frac{\lambda_n}{\alpha_n} \log \log \frac{1}{\pi_\infty^n(*)}} \right] e^{-m_n \lambda_n t} &= \left[ e^{-m_n \alpha_n T + 2 + \frac{\lambda_n}{\alpha_n} \log \log \frac{1}{\pi_\infty^n(*)}} \right] e^{-m_n \lambda_n (t-T)} \\ &\leq e^{B_2 + 2} e^{-A(t-T)} \\ &= B_1 e^{-At}. \end{aligned}$$

Then by (37), (40) holds. Thus by Lemma 3, (29) holds.  $\square$



**Remark 3.** When  $n$  tends to infinity, the discrete space is approximating a continuous space and thus the probability of any single point is tending to 0, that is,  $\pi_\infty^n(*)$  tends to 0. So  $\log \log \frac{1}{\pi_\infty^n(*)} \rightarrow \infty$  as  $n$  goes to infinity. Notice that by Lemma 15 in Appendix B,

$$\frac{\lambda_n}{\alpha_n} \leq \log(1/\pi_\infty^n(*)).$$

Therefore, in order to have

$$-m_n \lambda_n T + \frac{\lambda_n}{\alpha_n} \log \log \frac{1}{\pi_\infty^n(*)} \leq B_2,$$

it is sufficient to let

$$T m_n \lambda_n \geq O([\log(1/\pi_\infty^n(*)])^2). \quad (41)$$

Moreover, (41) can imply both (38) and (39).

According to the above analysis, our problem is simplified to checking that whether (41) is satisfied by the density dependent Markov chains  $\{X_n(t)\}$ .

By Remark 8 in Appendix B,  $\lambda_n$  is the smallest non-zero eigenvalue of  $I - \frac{K^n + K^{*,n}}{2}$ , where

$$K^n = \frac{Q^n}{m_n} + I$$

and  $K^{*,n}$  is adjoint to  $K^n$ . A matrix  $Q^{*,n}$  is said to be adjoint to the generator  $Q^n$ , if  $Q^n(\mathbf{x}, \mathbf{y})\pi_\infty^n(\mathbf{x})$  equals  $Q^{*,n}(\mathbf{y}, \mathbf{x})\pi_\infty^n(\mathbf{y})$ . Clearly,  $Q^{*,n} = m_n(K^{*,n} - I)$ , or equivalently,

$$K^{*,n} = \frac{Q^{*,n}}{m_n} + I.$$

So

$$m_n \left( I - \frac{K^n + K^{*,n}}{2} \right) = -\frac{Q^n + Q^{*,n}}{2}. \quad (42)$$

Denote the smallest non-zero eigenvalue of  $-\frac{Q^n + Q^{*,n}}{2}$  by  $\sigma_n$ . Then by (42),

$$m_n \lambda_n = \sigma_n. \quad (43)$$

Now we state our main result in this section.

**Theorem 4.** Let  $\{X_n(t)\}$  be the density dependent Markov chain derived from a given PEPA model. For each  $n \in \mathbb{N}$ , let  $S^n$  and  $\pi_\infty^n$  be the state space and steady-state probability distribution of  $X_n(t)$  respectively.  $Q^n$  is the infinitesimal generator of  $X_n(t)$  and  $\sigma_n$  is the smallest non-zero eigenvalue of  $-\frac{Q^n + Q^{*,n}}{2}$ , where  $Q^{*,n}$  is adjoint to  $Q^n$  in terms of  $\pi_\infty^n$ . If

$$\pi_\infty^n(*) \stackrel{\text{def}}{=} \min_{\mathbf{x} \in S^n} \pi_\infty^n(\mathbf{x}) \geq \frac{1}{\exp(O(\sqrt{\sigma_n}))} \quad (44)$$

for sufficiently large  $n$ , then  $X(t)$  has a finite limit as time tends to infinity, where  $X(t)$  is the solution of the corresponding derived ODEs from the same PEPA model.

*Proof.* By the given condition of  $\pi_\infty^n(*) \geq \frac{1}{\exp(O(\sqrt{\sigma_n}))}$ ,

$$\log \left[ \frac{1}{\pi_\infty^n(*)} \right] \leq \log[\exp(O(\sqrt{\sigma_n}))] = O(\sigma_n^{1/2}).$$

Thus

$$\left( \log \left[ \frac{1}{\pi_\infty^n(*)} \right] \right)^2 \leq O(\sigma_n).$$

Choose a large  $T$  such that

$$Tm_n\lambda_n = T\sigma_n \geq O(\sigma_n) \geq \left( \log \left[ \frac{1}{\pi_\infty^n(*)} \right] \right)^2.$$

By Remark 3 and Lemma 5, the particular condition holds. Therefore

$$\lim_{t \rightarrow \infty} X(t) = \hat{M}_\infty(\infty).$$

□

According to the above theorem, our problem is simplified to checking that whether (44) is satisfied by the density dependent Markov chains  $\{X_n(t)\}$ . In (44) both the spectral gap  $\lambda_n (= \frac{\sigma_n}{m_n})$  and  $\pi_\infty^n(*)$  are unknown. In fact, due to the state space explosion problem,  $\pi_\infty^n$  cannot be easily solved from  $\pi_\infty^n Q^n = 0$  or equivalently  $\pi_\infty^n K^n = \pi_\infty^n$ . Moreover, the estimation of the spectral gap for a given Markov chain in current literature (e.g. [32]) is heavily based on the known stationary distribution. Thus, the current results cannot provide a practical check for (44).

The convergence and consistency are supported by many numerical experiments, so we believe that (44) is unnecessary. In other words, we believe that (44) always holds in the context of PEPA, although at this moment we cannot prove it. Before concluding this section, we leave an open problem:

**Conjecture 1.** *The formula (44) in Theorem 4 holds for any PEPA model.*

## 6. Fluid analysis: an analytic approach (I)

The previous sections have demonstrated the fluid approximation and relevant analysis for PEPA. Some fundamental results about the derived ODEs such as the boundedness, nonnegativeness and convergence of the solutions, have been established through a probabilistic approach. In this section we will discuss the boundedness and nonnegativeness again, and prove them by a purely analytical argument. The convergence presented in the previous section is proved under a particular condition that cannot currently be easily checked. This section will present alternative approaches to deal with the convergence problem. In particular, for an interesting model with two synchronisations, its structural invariance as revealed in [22], will be shown to play an important role in the proof of the convergence. Moreover, for a class of PEPA models which have two component types and one synchronisation, an analytical proof of the convergence under some mild conditions on the populations will be presented. These discussions and investigations will provide a new insight into the fluid approximation of PEPA.

### 6.1. Analytical proof of boundedness and nonnegativeness

Recall that the set of derived ODEs from a general PEPA model is

$$\frac{d\mathbf{x}(U, t)}{dt} = - \sum_{\{l|U \in \text{pre}(l)\}} f(\mathbf{x}, l) + \sum_{\{l|U \in \text{post}(l)\}} f(\mathbf{x}, l). \quad (45)$$

As mentioned, in this formula the term  $\sum_{\{l|U \in \text{pre}(l)\}} f(\mathbf{x}, l)$  represents the exit rates in the local derivative  $U$ . An important fact to note is: the exit rates in a local derivative  $C_{i_j}$  in state  $\mathbf{x}$  are bounded by all the apparent rates in this local derivative. In fact, according to Proposition 1 in Section 2.2, if  $C_{i_j} \in \text{pre}(l)$  where  $l$  is a labelled activity, then the transition rate function  $f(\mathbf{x}, l)$  is bounded by  $r_l(\mathbf{x}, C_{i_j})$ , the apparent rates of  $l$  in  $C_{i_j}$  in state  $\mathbf{x}$ . That is,

$$f(\mathbf{x}, l) \leq r_l(\mathbf{x}, C_{i_j}) = \mathbf{x}[C_{i_j}]r_l(C_{i_j}). \quad (46)$$

We should point out that (46) is based on the discrete state space underlying the given model. According to our semantics of mapping PEPA model to ODEs, the fluid approximation-version of (46) also holds, i.e.  $f(\mathbf{x}(t), l) \leq \mathbf{x}(C_{i_j}, t)r_l(C_{i_j})$ . Hereafter the notation  $\mathbf{x}[\cdot]$  indicates a discrete state  $\mathbf{x}$ , while  $\mathbf{x}(\cdot)$  or  $\mathbf{x}(\cdot, t)$  reflects a continuous state  $\mathbf{x}$  at time  $t$ . Therefore, we have the following

**Proposition 7.** *For any local derivative  $C_{i_j}$ ,*

$$\sum_{\{l|C_{i_j} \in \text{pre}(l)\}} f(\mathbf{x}(t), l) \leq \mathbf{x}(C_{i_j}, t) \sum_{\{l|C_{i_j} \in \text{pre}(l)\}} r_l(C_{i_j}), \quad (47)$$

where  $r_l(C_{i_j})$  is the apparent rate of  $l$  in  $C_{i_j}$  for a single instance of  $C_{i_j}$  defined in Definition 5.

Proposition 7 and Propositions 3 can guarantee the boundedness and non-negativeness of the solutions. In the following, we will present an analytical proof of these properties, based on the two propositions. Suppose the initial values  $\mathbf{x}(C_{i_j}, 0)$  are given, and we denote  $\sum_j \mathbf{x}(C_{i_j}, 0)$  by  $N_{C_i}$ . We have a theorem:

**Theorem 5.** *If  $\mathbf{x}(C_{i_j}, t)$  satisfies (45) with nonnegative initial values, then*

$$0 \leq \mathbf{x}(C_{i_j}, t) \leq N_{C_i}, \quad \text{for any } t \geq 0. \quad (48)$$

Moreover, if the initial values are positive, then the solutions are always positive, i.e.,

$$0 < \mathbf{x}(C_{i_j}, t) < N_{C_i}, \quad \text{for any } t \geq 0. \quad (49)$$

*Proof.* By Proposition 3,  $\sum_j \mathbf{x}(C_{i_j}, t) = N_{C_i}$  for all  $t$ . All that is left to do is to prove that  $\mathbf{x}(C_{i_j}, t)$  is positive or nonnegative. The proof is divided into two cases.

Case 1: Suppose all the initial values are positive, i.e.  $\min_{i_j} \{\mathbf{x}(C_{i_j}, 0)\} > 0$ . We will show that  $\min_{i_j} \{\mathbf{x}(C_{i_j}, t)\} > 0$  for all  $t \geq 0$ . Otherwise, if there exists a  $t > 0$  such that  $\min_{i_j} \{\mathbf{x}(C_{i_j}, t)\} \leq 0$ , then there exists a point  $t' > 0$  such that  $\min_{i_j} \{\mathbf{x}(C_{i_j}, t')\} = 0$ . Let  $t^*$  be the first such point, i.e.

$$t^* = \inf \left\{ t > 0 \mid \min_{i_j} \{\mathbf{x}(C_{i_j}, t)\} = 0 \right\},$$

then  $0 < t^* < \infty$ . Without loss of generality, we assume  $\mathbf{x}(C_{1_1}, t)$  reaches zero at  $t^*$ , i.e.,

$$\mathbf{x}(C_{1_1}, t^*) = 0, \quad \mathbf{x}(C_{i_j}, t^*) \geq 0 \quad (i \neq 1 \vee j \neq 1)$$

and

$$\mathbf{x}(C_{i_j}, t) > 0, \quad t \in [0, t^*), \quad \forall i, j.$$

Thus, for  $t \in [0, t^*]$ , by Proposition 7,

$$\begin{aligned} \frac{d\mathbf{x}(C_{1_1}, t)}{dt} &= - \sum_{\{l|C_{1_1} \in \text{pre}(l)\}} f(\mathbf{x}, l) + \sum_{\{l|C_{1_1} \in \text{post}(l)\}} f(\mathbf{x}, l) \\ &\geq - \sum_{\{l|C_{1_1} \in \text{pre}(l)\}} f(\mathbf{x}, l) \\ &\geq -\mathbf{x}(C_{1_1}, t) \sum_{\{l|C_{1_1} \in \text{pre}(l)\}} r_l(C_{1_1}). \end{aligned}$$

Set  $R = \sum_{\{l|C_{1_1} \in \text{pre}(l)\}} r_l(C_{1_1})$ , then

$$\frac{d\mathbf{x}(C_{1_1}, t)}{dt} \geq -R\mathbf{x}(C_{1_1}, t). \quad (50)$$

By Lemma 14 in Appendix A, (50) implies

$$\mathbf{x}(C_{1_1}, t^*) \geq \mathbf{x}(C_{1_1}, 0)e^{-Rt^*} > 0.$$

This is a contradiction to  $\mathbf{x}(C_{1_1}, t^*) = 0$ . Therefore  $0 < \mathbf{x}(C_{i_j}, t)$ , and thus by Proposition 3,

$$0 < \mathbf{x}(C_{i_j}, t) < N_{C_0}, \quad \forall t.$$

Case 2: Suppose  $\min_{i_j} \{\mathbf{x}(C_{i_j}, 0)\} = 0$ . Let  $u_\delta(i_j, 0) = \mathbf{x}(C_{i_j}, 0) + \delta$  where  $\delta > 0$ . Let  $u_\delta(i_j, t)$  be the solution of (45), given the initial value  $u_\delta(i_j, 0)$ . By the proof of Case 1,  $u_\delta(i_j, t) > 0$  ( $\forall t \geq 0$ ). Noticing  $\min(\cdot)$  is a Lipschitz function, by the Fundamental Inequality in Appendix A, we have

$$|u_\delta(i_j, t) - \mathbf{x}(C_{i_j}, t)| \leq \delta e^{Kt}, \quad (51)$$

where  $K$  is a constant. Thus, for any given  $t \geq 0$ ,

$$\mathbf{x}(C_{i_j}, t) \geq u_\delta(i_j, t) - \delta e^{Kt} > -\delta e^{Kt}. \quad (52)$$

Let  $\delta \downarrow 0$  in (52), then we have  $\mathbf{x}(C_{i_j}, t) \geq 0$ . The proof is completed.  $\square$

## 6.2. A case study on convergence with two synchronisations

If a model has synchronisations, then the derived ODEs are nonlinear. The nonlinearity results in the complexity of the dynamic behaviour of fluid approximations. However, for some special models, we can still determine the convergence of the solutions. What follows is a case study for an interesting PEPA model, in which the structural property of invariance will be shown to play an important role in the proof of the convergence.

### 6.2.1. Theoretical study of convergence

The model considered here is given below,

**Model 2.**

$$\begin{array}{lll}
 X_1 & \stackrel{\text{def}}{=} & (\text{action1}, a_1).X_2 \\
 X_2 & \stackrel{\text{def}}{=} & (\text{action2}, a_2).X_1 \\
 Y_1 & \stackrel{\text{def}}{=} & (\text{action1}, a_1).Y_3 + (\text{job1}, c_1).Y_2 \\
 Y_2 & \stackrel{\text{def}}{=} & (\text{job2}, c_2).Y_1 \\
 Y_3 & \stackrel{\text{def}}{=} & (\text{job3}, c_3).Y_4 \\
 Y_4 & \stackrel{\text{def}}{=} & (\text{action2}, a_2).Y_2 + (\text{job4}, c_4).Y_3 \\
 (X_1[M_1] || X_2[M_2]) & \stackrel{\text{def}}{=} & (Y_1[N_1] || Y_2[N_2] || Y_3[N_3] || Y_4[N_4]) .
 \end{array}$$

$\bowtie_{\text{action1}, \text{action2}}$

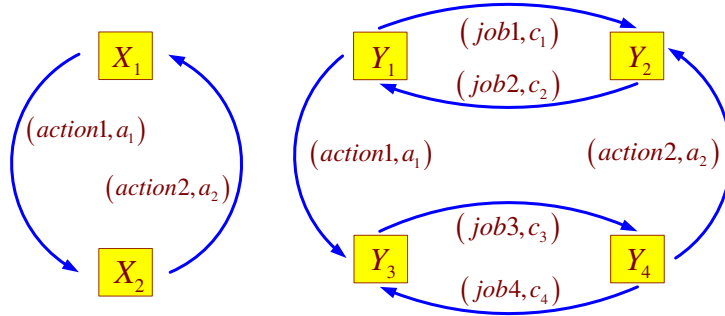


Figure 2: Transition systems of the components of Model 2

The operations of  $X$  and  $Y$  are illustrated in Figure 2. According to the mapping semantics, the derived ODEs from this model are

$$\left\{ \begin{array}{l} \frac{dx_1}{dt} = -a_1 \min\{x_1, y_1\} + a_2 \min\{x_2, y_4\} \\ \frac{dx_2}{dt} = a_1 \min\{x_1, y_1\} - a_2 \min\{x_2, y_4\} \\ \frac{dy_1}{dt} = -a_1 \min\{x_1, y_1\} - c_1 y_1 + c_2 y_2 \\ \frac{dy_2}{dt} = -c_2 y_2 + a_2 \min\{x_2, y_4\} + c_1 y_1 \\ \frac{dy_3}{dt} = -c_3 y_3 + c_4 y_4 + a_1 \min\{x_1, y_1\} \\ \frac{dy_4}{dt} = -a_2 \min\{x_2, y_4\} - c_4 y_4 + c_3 y_3 \end{array} \right. , \quad (53)$$

where  $x_i, y_j$  ( $i = 1, 2, j = 1, 2, \dots, 4$ ) denote the populations of  $X$  and  $Y$  in the local derivatives  $X_i, Y_j$  respectively. Now we state an interesting **assertion** for the specific PEPA model: the difference between the number of  $Y$  in their local derivatives  $Y_3$  and  $Y_4$ , and the number of  $X$  in the local derivative  $X_2$ , i.e.  $y_3 + y_4 - x_2$ , is a constant in any state. This fact can be explained as follows. Notice that there is only one way to increase  $y_3 + y_4$ , i.e. enabling the activity *action1*. As long as *action1* is activated, then there is a copy of  $Y$  entering the  $Y_3$  from  $Y_1$ . Meanwhile, since *action1* is shared by  $X$ , a corresponding copy of  $X$  will go to  $X_2$  from  $X_1$ . In other words,  $y_3 + y_4$  and  $x_2$  increase equally and simultaneously. On the other hand, there is also only one way to decrease  $y_3 + y_4$  and  $x_2$ , i.e. enabling the cooperated activity *action2*. This also allows  $y_3 + y_4$  and  $x_2$  to decrease both equally and simultaneously. So, the difference  $y_3 + y_4 - x_2$  will remain constant in any state and thus at any time. The assertion indicates that each state and therefore the whole state space of underlying CTMC may have some interesting structure properties, such as invariants. The techniques and applications of structural analysis of PEPA models have been developed in [22].

Throughout this section, let  $M$  and  $N$  be the total populations of the  $X$  and  $Y$  respectively, i.e.  $M = M_1 + M_2$  and  $N = N_1 + N_2 + N_3 + N_4$ . Notice  $y_1 + y_2 + y_3 + y_4 = N$  and  $x_1 + x_2 = M$  by the conservation law, so  $y_1 + y_2 - x_1 = N - M - (y_3 + y_4 - x_2)$  is another invariant because  $y_3 + y_4 - x_2$  is a constant. The fluid-approximation version of these two invariants also holds, which is illustrated by the following Lemma 6.

**Lemma 6.** For any  $t \geq 0$ ,

$$\begin{aligned} y_1(t) + y_2(t) - x_1(t) &= y_1(0) + y_2(0) - x_1(0), \\ y_3(t) + y_4(t) - x_2(t) &= y_3(0) + y_4(0) - x_2(0). \end{aligned}$$

*Proof.* According to (53), for any  $t \geq 0$ ,

$$\frac{d(y_1(t) + y_2(t) - x_1(t))}{dt} = \frac{dy_1}{dt} + \frac{dy_2}{dt} - \frac{dx_1}{dt} = 0.$$

So  $y_1(t) + y_2(t) - x_1(t) = y_1(0) + y_2(0) - x_1(0)$ ,  $\forall t \geq 0$ . By a similar argument, we also have  $y_3(t) + y_4(t) - x_2(t) = y_3(0) + y_4(0) - x_2(0)$ ,  $\forall t \geq 0$ .  $\square$

In the following we show how to use this kind of invariance to prove the convergence of the solution of (53) as time goes to infinity. Before presenting the results and the proof, we first rewrite (53) as follows:

$$\begin{aligned} \begin{pmatrix} \frac{dx_1}{dt} \\ \frac{dx_2}{dt} \\ \frac{dy_1}{dt} \\ \frac{dy_2}{dt} \\ \frac{dx_3}{dt} \\ \frac{dx_4}{dt} \end{pmatrix} &= I_{\{x_1 < y_1, x_2 < y_4\}} Q_1 \begin{pmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} + I_{\{x_1 < y_1, x_2 \geq y_4\}} Q_2 \begin{pmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} \\ &\quad + I_{\{x_1 \geq y_1, x_2 < y_4\}} Q_3 \begin{pmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} + I_{\{x_1 \geq y_1, x_2 \geq y_4\}} Q_4 \begin{pmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}, \end{aligned} \quad (54)$$

where the matrices  $Q_i$  ( $i = 1, 2, 3, 4$ ) are given as below:

$$Q_1 = \left( \begin{array}{cc|cc} -a_1 & a_2 & & \\ a_1 & -a_2 & & \\ \hline -a_1 & & -c_1 & c_2 \\ & a_2 & c_1 & -c_2 \\ a_1 & & & -c_3 & c_4 \\ & -a_2 & & c_3 & -c_4 \end{array} \right),$$



$$Q_2 = \left( \begin{array}{cc|cccc} -a_1 & 0 & & & & & a_2 \\ a_1 & 0 & & & & & -a_2 \\ \hline -a_1 & 0 & -c_1 & c_2 & & & \\ & 0 & c_1 & -c_2 & & & a_2 \\ a_1 & 0 & & & -c_3 & c_4 & \\ & 0 & & & c_3 & & -(c_4 + a_2) \end{array} \right),$$

$$Q_3 = \left( \begin{array}{cc|cccc} 0 & a_2 & & -a_1 & & & \\ 0 & -a_2 & & a_1 & & & \\ \hline 0 & & -a_1 - c_1 & c_2 & & & \\ 0 & a_2 & c_1 & -c_2 & & & \\ 0 & & a_1 & & -c_3 & c_4 & \\ 0 & -a_2 & & & c_3 & & -c_4 \end{array} \right),$$

$$Q_4 = \left( \begin{array}{cc|cccc} 0 & 0 & & -a_1 & & & a_2 \\ 0 & 0 & & a_1 & & & -a_2 \\ \hline 0 & 0 & -a_1 - c_1 & c_2 & & & \\ 0 & 0 & c_1 & -c_2 & & & a_2 \\ 0 & 0 & a_1 & & -c_3 & c_4 & \\ 0 & 0 & & & c_3 & & -a_2 - c_4 \end{array} \right).$$

As (54) illustrates, the derived ODEs are piecewise linear and they may be dominated by  $Q_i$  ( $i = 1, 2, 3, 4$ ) alternately. If the system is always dominated by only one specific matrix after a time, then the ODEs become linear after this time. For linear ODEs, as long as the eigenvalues of their coefficient matrices are either zeros or have negative real parts, then bounded solutions will converge as time tends to infinity, see Corollary 3 in Appendix D. Fortunately, here the eigenvalues of the matrices  $Q_i$  ( $i = 1, 2, 3, 4$ ) in Model 2 satisfy this property, the proof of which is shown in Appendix C. In addition, the solution of the derived ODEs from any PEPA model is bounded, as Theorem 5 illustrated. Therefore, if we can guarantee that after a time the ODEs (54) become linear, which means that one of the four matrices  $Q_i$  ( $i = 1, 2, 3, 4$ ) will be the coefficient matrix of the linear ODEs, then by Corollary 3 the solution will converge. So the convergence problem is reduced to determining whether the linearity can be finally guaranteed.

It is easy to see that the comparisons between  $x_1$  and  $y_1$ ,  $x_2$  and  $y_4$  determine the linearity. For instance, if after a time  $T$ , we always have  $x_1 > y_1$

and  $x_2 > y_4$ , then the matrix  $Q_4$  will dominate the system. Fortunately, the invariance in the model, as Lemma 6 reveals, can determine the comparisons in some circumstances. This is because this invariance reflects the relationship between different component types that are connected through synchronisations. This leads to several conclusions as follows.

**Proposition 8.** *If  $y_1(0) + y_2(0) \leq x_1(0)$  and  $y_3(0) + y_4(0) \leq x_2(0)$ , then the solution of (54) converges.*

*Proof.* By Lemma 6,  $y_1(t) + y_2(t) \leq x_1(t)$  and  $y_3(t) + y_4(t) \leq x_2(t)$  for all time  $t$ . Since both  $y_2(t)$  and  $y_4(t)$  are nonnegative by Theorem 5, we have  $y_1(t) \leq x_1(t)$  and  $y_4(t) \leq x_2(t)$  for any  $t$ . Thus, (54) becomes

$$\left( \frac{dx_1}{dt}, \frac{dx_2}{dt}, \frac{dy_1}{dt}, \frac{dy_2}{dt}, \frac{dy_3}{dt}, \frac{dy_4}{dt} \right)^T = Q_4(x_1, x_2, y_1, y_2, y_3, y_4)^T. \quad (55)$$

Notice that (55) is linear, and all eigenvalues of  $Q_4$  other than zeros have negative real parts, then according to Corollary 3, the solution of (55) converges as time goes to infinity.  $\square$

**Proposition 9.** *Suppose  $y_1(0) + y_2(0) > x_1(0)$  and  $y_3(0) + y_4(0) \leq x_2(0)$ . If either (I).  $N > \left(2 + \frac{a_1 + c_1}{c_2}\right)M$ , or (II).  $N > \frac{2(c_1 + c_2) + a_2}{c_2}M$ , where  $N > M > 0$  are the populations of  $Y$  and  $X$  respectively, then the solution of (54) converges.*

*Proof.* Suppose (I) holds. According to the conservation law,  $\sum_{i=1}^4 y_i(t) = N$ . By the boundedness of the solution, we have  $x_2(t) \leq M$ . Then by Lemma 6,  $y_3(t) + y_4(t) \leq x_2(t) \leq M$ . Therefore,

$$y_2(t) = N - (y_3(t) + y_4(t)) - y_1(t) \geq N - M - y_1(t). \quad (56)$$

Since  $\min\{x_1, y_1\} \leq y_1$ , so  $-a \min\{x_1, y_1\} \geq -a_1 y_1$ . Thus

$$\begin{aligned} \frac{dy_1}{dt} &= -a_1 \min\{x_1, y_1\} - c_1 y_1 + c_2 y_2 \\ &\geq -a_1 y_1 - c_1 y_1 + c_2 y_2 \\ &\geq -(a_1 + c_1) y_1 + c_2 (N - M - y_1) \\ &= -(a_1 + c_1 + c_2) y_1 + c_2 (N - M). \end{aligned} \quad (57)$$

That is

$$\frac{dy_1}{dt} \geq -(a_1 + c_1 + c_2)y_1 + c_2(N - M).$$

Applying Lemma 14 in Appendix A to this formula, we have

$$y_1(t) \geq \left( y_1(0) - \frac{c_2(N - M)}{a_1 + c_1 + c_2} \right) e^{-(a_1 + c_1 + c_2)t} + \frac{c_2(N - M)}{a_1 + c_1 + c_2}. \quad (58)$$

Since the first term of the right side of (58) converges to zero as time goes to infinity, i.e.  $\lim_{t \rightarrow \infty} \left( y_1(0) - \frac{c_2(N - M)}{a_1 + c_1 + c_2} \right) e^{-(a_1 + c_1 + c_2)t} = 0$ , and the second term  $\frac{c_2(N - M)}{a_1 + c_1 + c_2} > M$  which results from the condition  $N > \left( 2 + \frac{a_1 + c_1}{c_2} \right) M$ , then there exists  $T > 0$  such that for any  $t > T$ ,  $y_1(t) > M \geq x_1(t)$ . Then after time  $T$ , (54) becomes linear, and is dominated by  $Q_2$ . Because all eigenvalues of  $Q_2$  are either zeros or have negative real parts, the solution converges as time goes to infinity.

Now we assume (II) holds. Similarly, since  $\min\{x_2, y_4\} \leq x_2 \leq M$ , and  $y_1 \leq N - y_2$  which is due to  $y_1 + y_2 \leq N$ , we have

$$\begin{aligned} \frac{dy_2}{dt} &= -c_2y_2 + a_2 \min\{x_2, y_4\} + c_1y_1 \\ &\leq -c_2y_2 + a_2M + c_1y_1 \\ &\leq -c_2y_2 + a_2M + c_1(N - y_2) \\ &= -(c_1 + c_2)y_2 + a_2M + c_1N. \end{aligned} \quad (59)$$

By Lemma 14 in Appendix A,

$$y_2 \leq \left( y_2(0) - \frac{a_2M + c_1N}{c_1 + c_2} \right) e^{-(c_1 + c_2)t} + \frac{a_2M + c_1N}{c_1 + c_2}. \quad (60)$$

Therefore, since  $e^{-(c_1 + c_2)t}$  in above formula converges to zero as time tends to infinity, then for any  $\epsilon > 0$ , there exists  $T > 0$  such that for any time  $t > T$ ,

$$y_2 \leq \frac{a_2M + c_1N}{c_1 + c_2} + \epsilon. \quad (61)$$

Notice that the condition  $N > \frac{2(c_1 + c_2) + a_2}{c_2} M$  implies

$$\frac{c_2N - a_2M - (c_1 + c_2)M}{c_1 + c_2} > M,$$

and let  $\epsilon$  be small enough that  $\frac{c_2N - a_2M - (c_1 + c_2)M}{c_1 + c_2} - \epsilon > M$ . Then by (56),  $y_1 \geq (N - M) - y_2$ . Therefore,

$$\begin{aligned}
y_1 &\geq (N - M) - y_2 \\
&\geq (N - M) - \frac{a_2M + c_1N}{c_1 + c_2} - \epsilon \\
&= \frac{c_2N - a_2M - (c_1 + c_2)M}{c_1 + c_2} - \epsilon \\
&> M \geq x_1.
\end{aligned} \tag{62}$$

So  $y_1(t) > x_1(t)$ ,  $y_4(t) \leq x_2(t)$ , for any  $t > T$ , then by a similar argument the solution of (54) converges.  $\square$

Both condition (I) and (II) in Proposition 9 require  $N$  to be larger enough than  $M$ , to guarantee that  $y_1$  is larger than  $x_1$ . Since our PEPA model is symmetric, Proposition 9 has a corresponding symmetric version.

**Proposition 10.** *Suppose  $y_1(0) + y_2(0) \leq x_1(0)$  and  $y_3(0) + y_4(0) > x_2(0)$ . If either (I).  $N > \left(2 + \frac{a_2 + c_3}{c_4}\right)M$ , or (II).  $N > \frac{2(c_3 + c_4) + a_1}{c_1}M$ , where  $N > M > 0$  are the populations of  $Y$  and  $X$  respectively, then the solution of (54) converges.*

The proof of Proposition 10 is omitted here. We should point out that in our model the shared activity *action1* (respectively, *action2*) has the same local rate  $a_1$  (respectively,  $a_2$ ). We have taken the advantage of this in the above proofs. If the local rates of shared activities are not set to be the same, analogous conclusions can still hold but the discussion will be more complicated. However, the structural property invariance can still play an important role.

The above three propositions have illustrated the convergence for all situations in terms of the starting state, except for the case of  $y_1(0) + y_2(0) > x_1(0)$ ,  $y_3(0) + y_4(0) > x_2(0)$ . See a summary in Table 3. If  $y_1(0) + y_2(0) > x_1(0)$ ,  $y_3(0) + y_4(0) > x_2(0)$ , then the dynamic behaviour of the system is rather complex. A numerical study for this case will be presented in the next subsection.

Table 3: A summary for the convergence of Model 2

Starting state condition	Additional condition	Conclusion
$y_1(0) + y_2(0) \leq x_1(0),$ $y_3(0) + y_4(0) \leq x_2(0)$	No	Proposition 8
$y_1(0) + y_2(0) > x_1(0),$ $y_3(0) + y_4(0) \leq x_2(0)$	$N > k_1 M$	Proposition 9
$y_1(0) + y_2(0) \leq x_1(0),$ $y_3(0) + y_4(0) > x_2(0)$	$N > k_2 M$	Proposition 10
$y_1(0) + y_2(0) > x_1(0),$ $y_3(0) + y_4(0) > x_2(0)$	None identified	Explored numerically

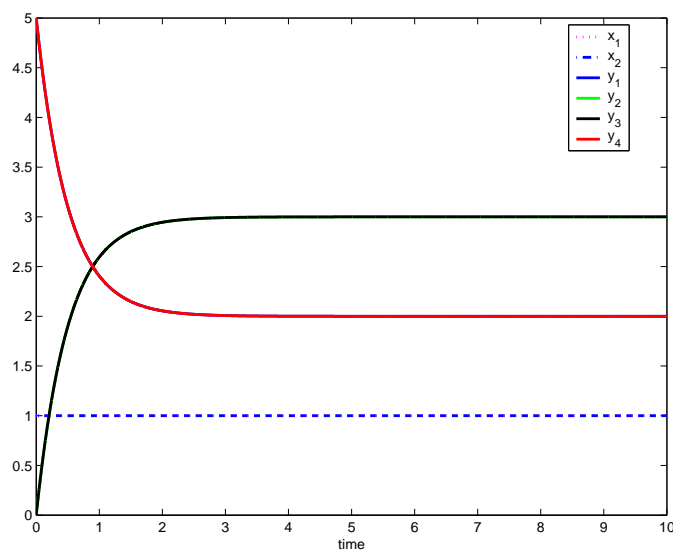


Figure 3: Numerical study for Model 2: rates  $(1, 1, 1, 1, 1, 1)$ ; equilibrium point  $(1, 1, 2, 3, 3, 2)$ . (Note: the curves of  $x_1$  and  $x_2$ , the curves of  $y_1$  and  $y_4$ , as well as those of  $y_2$  and  $y_3$  respectively, completely overlap.)

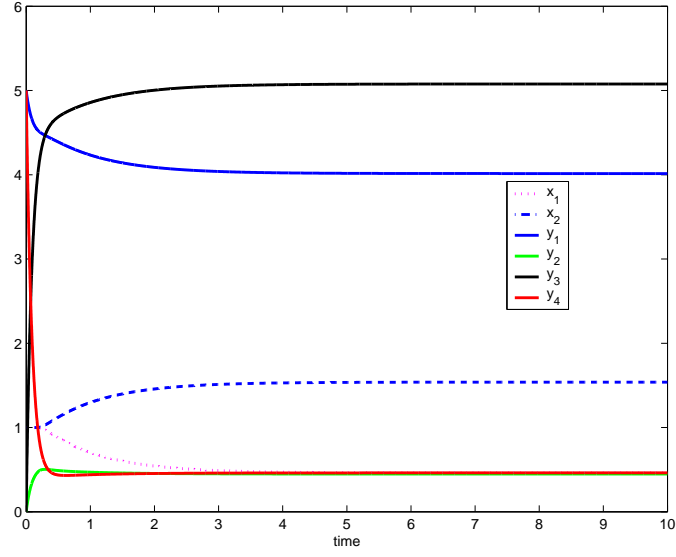


Figure 4: Numerical study for Model 2: rates  $(1, 1, 1, 10, 1, 10)$ ; equilibrium point  $(0.4616, 1.5384, 4.0140, 0.4476, 5.0769, 0.4615)$

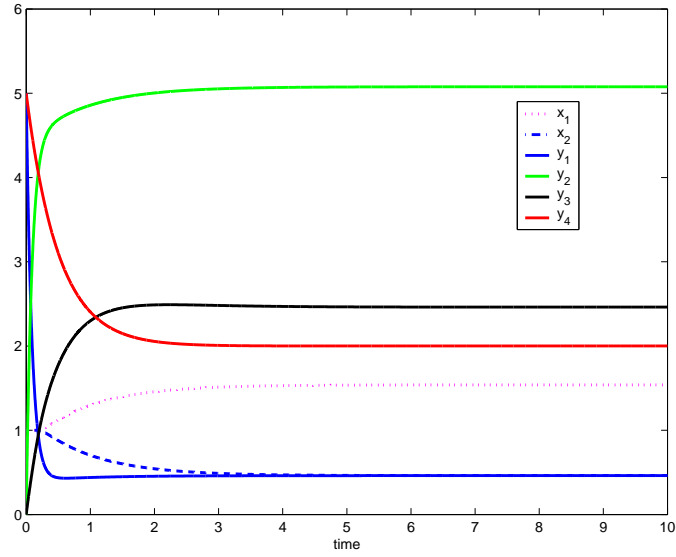


Figure 5: Numerical study for Model 2: rates  $(1, 1, 10, 1, 1, 1)$ ; equilibrium point  $(1.5384, 0.4616, 0.4615, 5.0769, 2.4616, 2.0000)$

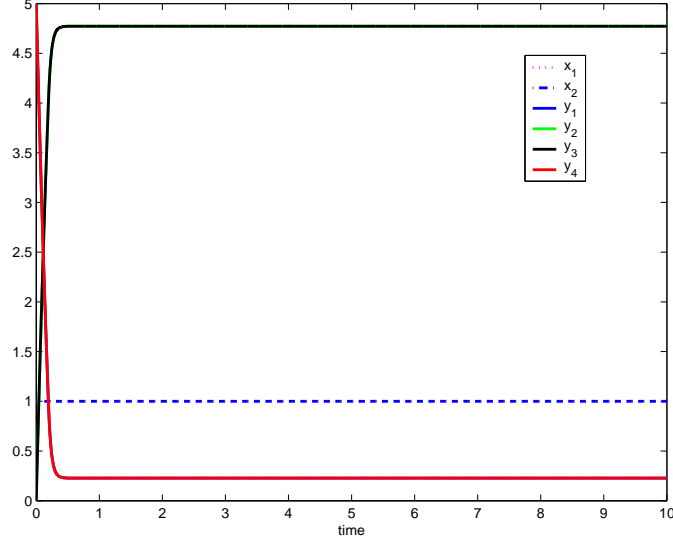


Figure 6: Numerical study for Model 2: rates  $(20, 20, 1, 1, 1, 1)$ ; equilibrium point  $(1, 1, 0.2273, 4.7727, 4.7727, 0.2273)$ . (Note: the curves of  $x_1$  and  $x_2$ , the curves of  $y_1$  and  $y_4$ , as well as those of  $y_2$  and  $y_3$  respectively, completely overlap.)

### 6.2.2. Experimental study of convergence

This subsection presents a numerical study at different action rate conditions. The starting state in this subsection is always assumed as  $(1, 1, 5, 0, 0, 5)$ , which satisfies the condition of  $y_1(0) + y_2(0) > x_1(0)$  and  $y_3(0) + y_4(0) > x_2(0)$ .

If all the action rates in the model are set to one, i.e.  $(a_1, a_2, c_1, c_2, c_3, c_4) = (1, 1, 1, 1, 1, 1)$ , then the equilibrium point of the ODEs is  $(x_1^*, x_2^*, y_1^*, y_2^*, y_3^*, y_4^*) = (1, 1, 2, 3, 3, 2)$ , as the numerical solution of the ODEs illustrates. In this case, the matrix  $Q_1$  finally dominates the system. See Figure 3. Notice that in this figure, the curves of  $x_1$  and  $x_2$  completely overlap, as well as the curves of  $y_1$  and  $y_4$ , and those of  $y_2$  and  $y_3$ .

In other situations, for example, if  $(a_1, a_2, c_1, c_2, c_3, c_4) = (1, 1, 1, 10, 1, 10)$  then the equilibrium point is  $(0.4616, 1.5384, 4.0140, 0.4476, 5.0769, 0.4615)$ , and the matrix  $Q_2$  eventually dominates the system. See Figure 4. Moreover, the matrices  $Q_3$  and  $Q_4$  can also finally dominate the system as long as the action rates are appropriately specified. See Figure 5 and Figure 6. We should point out that similarly to Figure 3, in Figure 6 the curves of  $x_1$  and  $x_2$ , the curves of  $y_1$  and  $y_4$ , as well as those of  $y_2$  and  $y_3$  respectively, completely overlap.

In short, the system dynamics is rather complex in the situation of  $y_1(0) + y_2(0) > x_1(0)$  and  $y_3(0) + y_4(0) > x_2(0)$ . A summary of these numerical studies is organised in Table 4.

Table 4: Complex dynamical behaviour of Model 2: starting state (1,1,5,0,0,5)

Rates: ( $a_1, a_2, c_1, c_2, c_3, c_4$ )	Equilibrium points: ( $x_1^*, x_2^*, y_1^*, y_2^*, y_3^*, y_4^*$ )	Dominator	Figure
(1, 1, 1, 1, 1, 1)	(1, 1, 2, 3, 3, 2)	$Q_1$	Figure 3
(1, 1, 1, 10, 1, 10)	$v_1^a$	$Q_2$	Figure 4
(1, 1, 10, 1, 1, 1)	$v_2^b$	$Q_3$	Figure 5
(20, 20, 1, 1, 1, 1)	(1, 1, 0.2273, 4.7727, 4.7727, 0.2273)	$Q_4$	Figure 6

$$^a v_1 = (0.4616, 1.5384, 4.0140, 0.4476, 5.0769, 0.4615)$$

$$^b v_2 = (1.5384, 0.4616, 0.4615, 5.0769, 2.4616, 2.0000)$$

## 7. Fluid analysis: an analytic approach (II)

### 7.1. Convergence for two component types and one synchronisation (I): a special case

The problem of convergence for more general models without strict conditions, is rather complex and has not been completely solved. But for a particular class of PEPA model — a model composed of two types of component with one synchronisation between them, we can determine the convergence of the solutions of the derived ODEs.

As discussed in the previous section, the ODEs derived from PEPA are piecewise linear and may be dominated by different coefficient matrices alternately. For any PEPA model which has two component types and one synchronisation, the two corresponding coefficient matrices can be proved to have a good property: their eigenvalues are either zeros or have negative real parts. The remaining issue for convergence is to ascertain that these two matrices will not always alternately dominate the system. In fact, we will prove that under some mild conditions, there exists a time after which there is only one coefficient matrix dominating the system. This means the ODEs become linear after that time. Since the coefficient matrix of the linear ODEs satisfies the good eigenvalue property, then by Corollary 3, the bounded solution will converge as time goes to infinity.



We first utilise an example in this section to show our approach to dealing with the convergence problem for this class of PEPA models. The proof for a general case in this class is presented in the next subsection.

#### 7.1.1. A previous model and the fluid approximation

Let us look at the following PEPA model, which is Model 1 presented previously:

$$\begin{aligned}
User_1 &\stackrel{def}{=}(task_1, a).User_2 \\
User_2 &\stackrel{def}{=}(task_2, b).User_1 \\
Provider_1 &\stackrel{def}{=}(task_1, a).Provider_2 \\
Provider_2 &\stackrel{def}{=}(reset, d).Provider_1 \\
(User_1[M]) &\boxtimes_{\{task_1\}} (Provider_1[N]).
\end{aligned}$$

The derived ODEs are as follows:

$$\begin{cases} \frac{dx_1}{dt} = -a \min\{x_1, y_1\} + bx_2 \\ \frac{dx_2}{dt} = a \min\{x_1, y_1\} - bx_2 \\ \frac{dy_1}{dt} = -a \min\{x_1, y_1\} + dy_2 \\ \frac{dy_2}{dt} = a \min\{x_1, y_1\} - dy_2 \end{cases} \quad (63)$$

where  $x_i$  and  $y_i$  represent the populations of  $User_i$  and  $Provider_i$  respectively,  $i = 1, 2$ . Clearly, (63) is equivalent to

$$\begin{pmatrix} \frac{dx_1}{dt} \\ \frac{dx_2}{dt} \\ \frac{dy_1}{dt} \\ \frac{dy_2}{dt} \end{pmatrix} = I_{\{x_1 \leq y_1\}} Q_1 \begin{pmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{pmatrix} + I_{\{x_1 > y_1\}} Q_2 \begin{pmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{pmatrix}, \quad (64)$$

where

$$Q_1 = \left( \begin{array}{cc|cc} -a & b & 0 & 0 \\ a & -b & 0 & 0 \\ -a & 0 & 0 & d \\ a & 0 & 0 & -d \end{array} \right), \quad Q_2 = \left( \begin{array}{cc|cc} 0 & b & -a & 0 \\ 0 & -b & a & 0 \\ 0 & 0 & -a & d \\ 0 & 0 & a & -d \end{array} \right). \quad (65)$$

Our interest is to see if the solution of (64) will converge as time goes to infinity. As we mentioned, this convergence problem can be divided into two subproblems, i.e. whether the nonlinear equations can finally become linear and whether the eigenvalues of the coefficient matrix are either zeros or have negative real parts. If answers to these two subproblems are both positive, then the convergence will hold.

The second subproblem can be easily dealt with. By calculations, the matrix  $Q_1$  has eigenvalues 0 (two folds),  $-d$ , and  $-(a+b)$ . Similarly,  $Q_2$  has eigenvalues 0 (two folds),  $-b$ ,  $-(a+d)$ . Therefore, the eigenvalues of  $Q_1$  and  $Q_2$  other than zeros are negative. Moreover, for a general PEPA model which has two component types and one synchronisation, Theorem 7 in the next section shows that the corresponding coefficient matrices always have this property.

The remaining work to determine the convergence of the ODE solution, is to solve the first subproblem, i.e. to ascertain that after a time it is always the case that  $x_1 > y_1$  or  $x_1 \leq y_1$ . In this model, there is no invariance relating the two different component types, so we cannot rely on invariants to investigate this subproblem. However, we have a new way to cope with this problem.

#### 7.1.2. Proof outline of convergence

Notice that  $y_1(t) \leq N$  by the boundedness of solutions. If we can prove that after time  $T$ ,  $x_1(t) \geq cM$ , where  $c > 0$  is independent of  $M$ , we will get, provided  $cM > N$ ,

$$x_1(t) \geq cM > N \geq y_1(t), \quad t \geq T.$$

Therefore, the ODEs (63) will become linear after time  $T$ . In the following, we identify that  $x_1(t) \geq cM$ .

Let

$$\alpha(t) = \begin{cases} \frac{\min\{x_1(t), y_1(t)\}}{x_1(t)}, & x_1(t) \neq 0, \\ 1, & x_1(t) = 0, \end{cases}$$

then  $0 \leq \alpha(t) \leq 1$  by the nonnegativity of  $x_1(t)$  and  $y_1(t)$ . The ODEs associated with component type  $X$  can be rewritten as

$$\begin{cases} \frac{dx_1}{dt} = -a\alpha(t)x_1 + bx_2, \\ \frac{dx_2}{dt} = a\alpha(t)x_1 - bx_2. \end{cases} \quad (66)$$

Let

$$A(t) = \begin{pmatrix} -a\alpha(t) & b \\ a\alpha(t) & -b \end{pmatrix}, \quad X(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}.$$

Then (66) can be written as

$$\frac{dX(t)}{dt} = A(t)X(t). \quad (67)$$

The solution of (67) is

$$X(t) = e^{\int_0^t A(s)ds} X(0). \quad (68)$$

Let  $B(t) = \frac{1}{t} \int_0^t A(s)ds$ , then

$$X(t) = e^{\int_0^t A(s)ds} X(0) = e^{tB(t)} X(0), \quad (69)$$

and

$$B(t) = \frac{1}{t} \begin{pmatrix} -a \int_0^t \alpha(s)ds & bt \\ a \int_0^t \alpha(s)ds & -bt \end{pmatrix} = \begin{pmatrix} -a\beta(t) & b \\ a\beta(t) & -b \end{pmatrix}, \quad (70)$$

where  $\beta(t) = \frac{\int_0^t \alpha(s)ds}{t}$ . Obviously,  $0 \leq \beta(t) \leq 1$  because  $0 \leq \alpha(s) \leq 1$ . Notice that the matrix  $B(t)$  can be diagonalised as

$$\begin{aligned} B(t) &= \begin{pmatrix} -a\beta(t) & b \\ a\beta(t) & -b \end{pmatrix} \\ &= \begin{pmatrix} \frac{b}{a\beta(t)+b} & 1 \\ \frac{a\beta(t)}{a\beta(t)+b} & -1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & -(a\beta(t)+b) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \frac{a\beta(t)}{a\beta(t)+b} & -\frac{b}{a\beta(t)+b} \end{pmatrix} \\ &= U(t) \begin{pmatrix} 0 & 0 \\ 0 & -(a\beta(t)+b) \end{pmatrix} U^{-1}(t), \end{aligned} \quad (71)$$

where

$$U(t) = \begin{pmatrix} \frac{b}{a\beta(t)+b} & 1 \\ \frac{a\beta(t)}{a\beta(t)+b} & -1 \end{pmatrix}, \quad U^{-1}(t) = \begin{pmatrix} 1 & 1 \\ \frac{a\beta(t)}{a\beta(t)+b} & -\frac{b}{a\beta(t)+b} \end{pmatrix}, \quad (72)$$

so

$$X(t) = e^{tB(t)} X(0) = U(t) \begin{pmatrix} 1 & 0 \\ 0 & e^{-t(a\beta(t)+b)} \end{pmatrix} U^{-1}(t) X(0). \quad (73)$$

In the formula (71), both 0 and  $-(a\beta(t) + b)$  are  $B(t)$ 's eigenvalues, while the columns of  $U(t)$  are the corresponding eigenvectors. In particular,  $\left(\frac{b}{a\beta(t) + b}, \frac{a\beta(t)}{a\beta(t) + b}\right)^T$ , i.e. the first column of  $U(t)$ , is the eigenvector corresponding to the eigenvalue zero.

We define a function  $\hat{X}(t)$  by

$$\hat{X}(t) = U(t) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} U^{-1}(t) X(0). \quad (74)$$

By simple calculation,

$$\hat{X}(t) = (\hat{x}_1(t), \hat{x}_2(t))^T = \left(\frac{bM}{a\beta(t) + b}, \frac{a\beta(t)M}{a\beta(t) + b}\right)^T. \quad (75)$$

Clearly,  $\hat{X}(t)$  is the normalised eigenvector corresponding the zero eigenvalue (at the time  $t$ ). Here the normalisation is in terms of the total population  $M$  of the component type  $X$ . Moreover,  $\hat{X}(t)$  embodies some equilibrium meaning. In fact, we have a conclusion:

$$\lim_{t \rightarrow \infty} \|X(t) - \hat{X}(t)\| = 0. \quad (76)$$

Since the explicit expression of  $U(t)$  is available, the proof of (76) is easy and thus omitted. This conclusion is also included in Proposition 12, the proof of which does not rely on the explicit information of  $U(t)$  and will be introduced later.

Now we discuss the benefit brought by this formula. By (76) the first entry of  $X(t)$  approximates the first entry of  $\hat{X}(t)$ , i.e.  $x_1(t)$  approximates  $\hat{x}_1(t) = \frac{bM}{a\beta(t) + b}$ . Thus, for any  $\epsilon > 0$ , there exists  $T > 0$  such that for any  $t \geq T$ ,

$$x_1(t) > \hat{x}_1(t) - \epsilon = \frac{bM}{a\beta(t) + b} - \epsilon.$$

Since  $\frac{bM}{a+b} \leq \frac{bM}{a\beta(t) + b} \leq M$  because  $0 \leq \beta(t) \leq 1$ , so  $x_1(t) > \frac{bM}{a+b} - \epsilon$ .

Therefore, if  $\frac{bM}{a+b} > N$ , then by the boundedness of  $y_1(t)$ , i.e.  $y_1(t) \leq N$ , we have

$$x_1(t) > \frac{bM}{a+b} - \epsilon > N \geq y_1(t),$$

as long as  $\epsilon$  is small enough. This means that  $Q_2$  will dominate the system after time  $T$ . So we have

**Proposition 11.** *If  $\frac{bM}{a+b} > N$ , then the solution of the ODEs (64) converges as time tends to infinity.*

Since the model is symmetric, this proposition has a symmetric version: if  $\frac{dN}{a+d} > M$ , then the solution of the ODEs also converges.

As we discussed, there are two key steps in the proof of Proposition 11. The first step is to establish the approximation of  $x_1(t)$  to  $\hat{x}_1(t)$ , i.e.  $x_1(t) \approx \hat{x}_1(t)$ . The second one is to give an estimation  $\hat{x}_1(t) \geq cM$ . According to these two conclusions, we have  $x_1(t) \geq c'M$  where  $c' < c$ , and therefore can conclude that  $x_1(t) \geq c'M > N > y_1(t)$  provided the condition  $c'M > N$ . This is the main philosophy behind our proof for  $x_1(t) > y_1(t)$ .

For the sake of generality, the proofs of these two conclusions should not rely on the explicit expressions of the eigenvalues and eigenvectors of  $B(t)$ . This is because for general PEPA models with two component types and one synchronisation, these explicit expressions are not always available. The following subsection will present our discussions about these steps, and the proofs for the conclusions which do not rely on these explicit expressions.

### 7.1.3. Proof not relying on explicit expressions

This subsection will divide into two parts. In the first part, we will give a lower bound for the eigenvalues of the coefficient matrix  $B(t)$ , based on which a proof of the approximation of  $X(t)$  to  $\hat{X}(t)$  is given. The second part will establish the estimation  $\hat{x}_1(t) \geq cM$ . All proofs in this subsection do not require knowledge of the explicit expressions of the eigenvalues and eigenvectors of  $B(t)$ .

For convenience, in this subsection we define

$$f(\beta) = \begin{pmatrix} -a\beta & b \\ a\beta & -b \end{pmatrix}, \quad (77)$$

where  $f$  is a matrix-valued function defined on  $\mathbb{R}$ . Then the matrix

$$B(t) = \begin{pmatrix} -a\beta(t) & b \\ a\beta(t) & -b \end{pmatrix}$$

can be written as  $B(t) = f(\beta(t))$ . The diagonalisation of  $f(\beta)$  is

$$f(\beta) = g(\beta) \begin{pmatrix} 0 & 0 \\ 0 & \lambda(\beta) \end{pmatrix} g^{-1}(\beta),$$

where  $\lambda(\beta)$  is  $f(\beta)$ 's nonzero eigenvalue, and  $g(\beta)$  is a matrix whose columns are the eigenvectors of  $f(\beta)$ . Here  $g^{-1}(\beta)$  is the inverse of the matrix  $g(\beta)$ . Notice that  $\lambda(\beta)$  is real, because if  $\lambda(\beta)$  is complex then its conjugation must be an eigenvalue, which is contradicted by the fact that  $f(\beta)$  only has two eigenvalues, 0 and  $\lambda(\beta)$ . The following discussions in this subsection do not rely on the explicit expressions of  $\lambda(\beta)$ ,  $g(\beta)$  and  $g^{-1}(\beta)$ , although it is easy to see that  $\lambda(\beta) = -a\beta + b$  and

$$g(\beta) = \begin{pmatrix} \frac{b}{a\beta+b} & 1 \\ \frac{a\beta}{a\beta+b} & -1 \end{pmatrix}, \quad g^{-1}(\beta) = \begin{pmatrix} 1 & 1 \\ \frac{a\beta}{a\beta+b} & -\frac{b}{a\beta+b} \end{pmatrix}.$$

### 1. $X(t)$ approximates $\hat{X}(t)$

In the following, we will give a lower bound for the nonzero eigenvalue of  $f(\beta)$ , i.e.  $\lambda(\beta)$ , and based on this prove the approximation of  $A(t)$  to  $\hat{X}(t)$  as time tends to infinity.

If  $\beta > 0$ , then the transpose of  $f(\beta)$ , i.e.  $f(\beta)^T$ , is an infinitesimal generator, and thus the nonzero eigenvalue  $\lambda(\beta)$  has negative real part, i.e.  $\Re(\lambda(\beta)) < 0$ . If  $\beta = 0$ , then  $f(\beta)$  is independent of  $\beta$  and becomes a nonnegative matrix, i.e. each entry of it is nonnegative. Based on the Perron-Frobenius theorem which is presented in the next subsection, we can still have  $\Re(\lambda(0)) < 0$ . Therefore, for any  $\beta$ ,  $f(\beta)$ 's eigenvalue other than zero has negative real part. This conclusion is stated in the following lemma.

**Lemma 7.** *For any  $\beta \in [0, 1]$ ,  $\Re(\lambda(\beta)) < 0$ , where  $\lambda(\beta)$  is a nonzero eigenvalue of  $f(\beta)$ .*

The proof of Lemma 7 is presented in Appendix D.3. Lemma 7 can further lead to the following

**Lemma 8.** *Let*

$$\Lambda_1 = \inf_{\beta \in [0, 1]} \{-\Re(\lambda(\beta)) \mid \lambda(\beta) \text{ is } f(\beta) \text{'s non-zero eigenvalue}\}, \quad (78)$$

*then  $\Lambda_1 > 0$ .*

*Proof.* By Lemma 7,  $-\Re(\lambda(\beta)) > 0$  for any  $\beta \in [0, 1]$ , so  $\Lambda_1 \geq 0$ . Suppose  $\Lambda_1 = 0$ . Because the eigenvalue  $\lambda(\beta)$  is a continuous function of the matrix  $f(\beta)$ , where  $f(\beta)$  is also continuous on  $[0, 1]$  with respect to  $\beta$ , so  $\lambda(\beta)$  is a continuous function of  $\beta$  on  $[0, 1]$ . This is due to the fact that a composition of continuous functions is still continuous. Noticing  $\Re(\cdot)$  is also a continuous function, so  $-\Re(\lambda(\beta))$  is continuous with respect to  $\lambda(\beta)$ , and thus with respect to  $\beta$  on  $[0, 1]$ . Since a continuous function on a closed interval can achieve its minimum, there exists  $\beta_0 \in [0, 1]$  such that  $-\Re(\lambda(\beta_0))$  achieves the minimum  $\Lambda_1$ , i.e.  $-\Re(\lambda(\beta_0)) = \Lambda_1 = 0$ . This is contradicted to Lemma 7. Therefore,  $\Lambda_1 > 0$ .  $\square$

For any  $t \in [0, \infty)$ ,  $\beta(t) \in [0, 1]$ . Noticing  $B(t) = f(\beta(t))$ , therefore

$$\begin{aligned} & \{\lambda \mid \lambda \text{ is } B(t)\text{'s non-zero eigenvalue}, t > 0\} \\ &= \{\lambda \mid \lambda \text{ is } f(\beta(t))\text{'s non-zero eigenvalue}, t > 0\} \\ &\subseteq \{\lambda \mid \lambda \text{ is } f(\beta)\text{'s non-zero eigenvalue}; \beta \in [0, 1]\}. \end{aligned} \quad (79)$$

Thus,

$$\begin{aligned} \Lambda &\triangleq \inf\{-\Re(\lambda) \mid \lambda \text{ is } B(t)\text{'s nonzero eigenvalue}, t > 0\} \\ &\geq \inf\{-\Re(\lambda) \mid \lambda \text{ is } f(\beta)\text{'s non-zero eigenvalue}; \beta \in [0, 1]\} = \Lambda_1. \end{aligned} \quad (80)$$

Because  $\Lambda_1 > 0$  by Lemma 8, so  $\Lambda > 0$ . That is,

**Corollary 1.** *Let*

$$\Lambda = \inf_{t \geq 0} \{-\Re(\lambda(t)) \mid \lambda(t) \text{ is } B(t)\text{'s non-zero eigenvalue}\},$$

*then  $\Lambda > 0$ .*

Based on this corollary, we can prove the approximation of  $X(t)$  to  $\hat{X}(t)$ .

**Proposition 12.** *Let  $X(t) = (x_1(t), x_2(t))^T = e^{tB(t)}X(0)$ , i.e. the solution of (66). Let  $\hat{X}(t)$  be defined by (74), i.e.*

$$\hat{X}(t) = \begin{pmatrix} \hat{x}_1(t) \\ \hat{x}_2(t) \end{pmatrix} = U(t) \begin{pmatrix} 1 & \\ & 0 \end{pmatrix} U(t)^{-1} \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix}. \quad (81)$$

*Then  $\lim_{t \rightarrow \infty} \|X(t) - \hat{X}(t)\| = 0$ .*

*Proof.* Notice that eigenvectors of a matrix are continuous functions of the matrix. Since  $g(\beta)$  is composed of the eigenvectors of the matrix  $f(\beta)$  and  $f(\beta)$  is continuous on  $[0, 1]$  with respect to  $\beta$ , therefore  $g(\beta)$  is continuous on  $[0, 1]$  with respect to  $\beta$ . Because the inverse of a matrix is a continuous mapping, so  $g^{-1}(\beta)$ , i.e. the inverse of  $g(\beta)$ , is continuous with respect to  $g(\beta)$ , and therefore is continuous on  $[0, 1]$  with respect to  $\beta$  since  $g(\beta)$  is continuous on  $[0, 1]$ . Since any continuous function is bounded on a compact set  $[0, 1]$ , both  $g(\beta)$  and  $g^{-1}(\beta)$  are bounded on  $[0, 1]$ . That is, there exists  $K > 0$  such that  $\|g(\beta)\| \leq K$  and  $\|g^{-1}(\beta)\| \leq K$  for all  $\beta \in [0, 1]$ . Because

$$\{U(t) \mid t \in [0, \infty)\} = \{g(\beta(t)) \mid t \in [0, \infty)\} \subseteq \{g(\beta) \mid \beta \in [0, 1]\},$$

we have

$$\sup_{t \geq 0} \|U(t)\| \leq \sup_{\beta \in [0, 1]} \|g(\beta)\| \leq K.$$

Similarly,  $\sup_{t \geq 0} \|U^{-1}(t)\| \leq K$ . Notice

$$\begin{aligned} X(t) - \hat{X}(t) &= e^{tB(t)} X(0) - U(t) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} U(t)^{-1} X(0) \\ &= \left[ U(t) \begin{pmatrix} 1 & 0 \\ 0 & e^{t\lambda(t)} \end{pmatrix} U(t)^{-1} - U(t) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} U(t)^{-1} \right] X(0) \\ &= U(t) \begin{pmatrix} 0 & 0 \\ 0 & e^{t\lambda(t)} \end{pmatrix} U(t)^{-1} X(0), \end{aligned}$$

where  $\lambda(t)$  is  $B(t)$ 's nonzero eigenvalue. By a similar argument to  $\lambda(\beta)$ ,  $\lambda(t)$  is also real. Therefore,

$$-\lambda(t) = \Re(-\lambda(t)) = -\Re(\lambda(t)) \geq \Lambda > 0$$

or  $\lambda(t) \leq -\Lambda < 0$ , where  $\Lambda$  is defined in Corollary 1. Then

$$\begin{aligned} \|X(t) - \hat{X}(t)\| &= \left\| U(t) \begin{pmatrix} 0 & 0 \\ 0 & e^{t\lambda(t)} \end{pmatrix} U(t)^{-1} X(0) \right\| \\ &\leq \|U(t)\| \left\| \begin{pmatrix} 0 & 0 \\ 0 & e^{t\lambda(t)} \end{pmatrix} \right\| \|U(t)^{-1}\| \|X(0)\| \\ &\leq K^2 \|X(0)\| e^{t\lambda(t)} \\ &\leq K^2 \|X(0)\| e^{-t\Lambda}. \end{aligned}$$

Here we have used the norm property:  $\|AB\| \leq \|A\| \|B\|$ . Since  $\Lambda > 0$ , we have  $\lim_{t \rightarrow \infty} \|X(t) - \hat{X}(t)\| = 0$ .  $\square$



## 2: An lower-bound estimation on population in local derivatives

In the following, we will prove that there exists  $T$ , such that  $\hat{x}_1(t) \geq cM$  for any  $t > T$ . We first define a function

$$h(\beta) = (h_1(\beta), h_2(\beta))^T = g(\beta) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} g^{-1}(\beta) \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix}.$$

Clearly, we have  $\hat{X}(t) = h(\beta(t))$  and  $\hat{x}_1(t) = h_1(\beta(t))$ . Since  $\beta(t) \in [0, 1]$  for all  $t$ , the following proposition can imply  $\hat{x}_1(t) \geq cM$ .

**Proposition 13.** *There exists  $c > 0$  such that*

$$\inf_{\beta \in [0, 1]} h_1(\beta) \geq cM.$$

where  $M = x_1(0) + x_2(0)$ ,  $c$  is independent of  $M$ .

*Proof.* Without loss of generality, we assume  $M = 1$ . We will show  $\inf_{\beta \in [0, 1]} h_1(\beta) = c > 0$ . Since  $h_1(\beta)$  is a continuous function of  $\beta$  which is due to the continuity of  $g(\beta)$  and  $g^{-1}(\beta)$ ,  $h_1(\beta)$  can achieve its minimum on  $[0, 1]$ . That is, there exists  $\beta_0 \in [0, 1]$ , such that

$$h_1(\beta_0) = \inf_{\beta \in [0, 1]} h_1(\beta) = c.$$

Consider the matrix

$$f(\beta_0) = \begin{pmatrix} -a\beta_0 & b \\ a\beta_0 & -b \end{pmatrix}$$

and a set of linear ODEs

$$\begin{pmatrix} \frac{dz_1}{dt} \\ \frac{dz_2}{dt} \end{pmatrix} = f(\beta_0) \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}. \quad (82)$$

The solution of (82), given an initial value  $Z(0) = X(0) = (x_1(0), x_2(0))^T$ , is  $Z(t) = e^{tf(\beta_0)} X(0)$ .

According to (71),  $f(\beta_0)$  can be diagonalised as

$$f(\beta_0) = g(\beta_0) \begin{pmatrix} 0 & 0 \\ 0 & \lambda(\beta_0) \end{pmatrix} g^{-1}(\beta_0).$$

where  $\lambda(\beta_0)$  is the nonzero and real eigenvalue of  $f(\beta_0)$ . Thus

$$Z(t) = e^{tf(\beta_0)} X(0) = g(\beta_0) \begin{pmatrix} 1 & 0 \\ 0 & e^{t\lambda(\beta_0)} \end{pmatrix} g(\beta_0)^{-1} \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix}. \quad (83)$$

Because  $\lambda(\beta_0) < 0$  by Lemma 7, so as time goes to infinity,

$$\begin{aligned} Z(t) &= g(\beta_0) \begin{pmatrix} 1 & 0 \\ 0 & e^{t\lambda(\beta_0)} \end{pmatrix} g(\beta_0)^{-1} \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix} \\ &\longrightarrow g(\beta_0) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} g(\beta_0)^{-1} \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix} = h(\beta_0). \end{aligned} \quad (84)$$

That is,  $\lim_{t \rightarrow \infty} Z(t) = h(\beta_0)$ . In the following, we discuss two possible cases:  $\beta_0 > 0$  and  $\beta_0 = 0$ .

If  $\beta_0 > 0$ , then the transpose of the matrix  $f(\beta_0)$ , i.e.  $f(\beta_0)^T$ , is an infinitesimal generator of an irreducible CTMC, which has two states and the transition rates between these two states are  $a\beta_0$  and  $b$  respectively. Moreover, the transient distribution of this CTMC, denoted by  $Z(t) = (z_1(t), z_2(t))^T$ , satisfies the ODEs (82). As time goes to infinity, the transient distribution  $Z(t)$  converges to the unique steady-state probability distribution. Since  $\lim_{t \rightarrow \infty} Z(t) = h(\beta_0)$ , therefore  $h(\beta_0) = (h_1(\beta_0), h_2(\beta_0))^T$  is the steady-state probability distribution and thus  $h_1(\beta_0) > 0$ . So  $\inf_{\beta \in [0,1]} h_1(\beta) = h_1(\beta_0) > 0$ .

If  $\beta_0 = 0$ , then

$$f(\beta_0) = \begin{pmatrix} 0 & b \\ 0 & -b \end{pmatrix}.$$

Since  $\lim_{t \rightarrow \infty} Z(t) = h(\beta_0)$ , therefore  $\left( \frac{dz_1}{dt}, \frac{dz_2}{dt} \right)^T$  converges to zero. Letting time go to infinity on the both sides of (82), we obtain the following equilibrium equations,

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = f(\beta_0) \begin{pmatrix} h_1(\beta_0) \\ h_2(\beta_0) \end{pmatrix}. \quad (85)$$

By the conservation law,  $(h_1(\beta_0) + h_2(\beta_0))^T = M = 1$ . Therefore,  $(h_1(\beta_0), h_2(\beta_0))^T$  satisfies

$$\begin{cases} f(\beta_0)(h_1(\beta_0), h_2(\beta_0))^T = 0, \\ h_1(\beta_0) + h_2(\beta_0) = M = 1. \end{cases} \quad (86)$$

Solving (86), we obtain the unique solution  $(h_1(\beta_0), h_2(\beta_0))^T = (1, 0)^T$ . Therefore,  $h_1(\beta_0)$  is one, and thus  $\inf_{\beta \in [0,1]} h_1(\beta) = h_1(\beta_0) > 0$ .  $\square$

**Remark 4.** As  $\beta$  tends to 0,

$$f(\beta) = \begin{pmatrix} -a\beta & b \\ a\beta & -b \end{pmatrix} \longrightarrow f(0) = \begin{pmatrix} 0 & b \\ 0 & -b \end{pmatrix}.$$

Correspondingly, for the equilibrium  $h(\beta) = (h_1(\beta), h_2(\beta))^T = \left( \frac{bM}{a\beta+b}, \frac{a\beta M}{a\beta+b} \right)^T$  satisfying  $f(\beta)h(\beta) = 0$  and  $h_1(\beta) + h_2(\beta) = M$ , we have  $\left( \frac{bM}{a\beta+b}, \frac{a\beta M}{a\beta+b} \right)^T \rightarrow (M, 0)^T$  as  $\beta$  tends to zero. From the explicit expression, i.e.  $h_1(\beta) = \frac{bM}{a\beta+b}$ , the minimum and maximum of  $h_1(\beta)$  are  $\frac{bM}{a+b}$  and  $M$  respectively, which correspond to the matrices  $f(1)$  and  $f(0)$  respectively. In the context of the PEPA model,  $f(1)$  corresponds to a free subsystem and there is no synchronisation effect on it, i.e. the subsystem of component type  $X$  is independent of  $Y$ . The matrix  $f(0)$  reflects that the subsystem of  $X$  has been influenced by the subsystem of  $Y$ , i.e. the rates of shared activities are determined by  $Y$ , that is, the term  $a \min\{x_1, y_1\}$  has been replaced by  $ay_1$ . Therefore the exit rates from the local derivative  $X_1$  correspondingly become smaller since now  $ay_1 < ax_1$ . In order to balance the flux, which is described by the equilibrium equation, the population of  $X_1$  must increase. That is why the equilibrium  $h_1(\beta)$  increases as  $\beta$  decreases. In short, synchronisations can increase the populations in synchronised local derivatives in the steady state.

As an application of the above facts, if  $h_1(\beta_0) > 0$  for some  $\beta_0 > 0$ , then we can claim that  $h_1(0) > 0$  because  $h_1(0) \geq h_1(\beta_0) > 0$ .

Obviously, Proposition 13 has a corollary:

**Corollary 2.** There exists  $c > 0$  such that for any  $t \in [0, \infty)$ ,  $\hat{x}_1(t) \geq cM$ .

Proposition 12 and Proposition 13 can lead to the following lemma.

**Lemma 9.** There exists  $c > 0, T > 0$ , such that  $x_1(t) \geq cM$  for all  $t \geq T$ .

*Proof.* By Proposition 13 or Corollary 2, there exists  $c_1, T_1 > 0$  such that  $\hat{x}_1(t) \geq c_1M$  for any  $t > T_1$ . By Proposition 12,  $\lim_{t \rightarrow \infty} |x_1(t) - \hat{x}_1(t)| = 0$ , which implies that for any  $\epsilon$ , there exists  $T_2 > 0$  such that for any  $t > T_2$ ,  $x_1(t) > \hat{x}_1(t) - \epsilon$ . Choose  $T_2 > T_1$ , then we have

$$x_1(t) > \hat{x}_1(t) - \epsilon \geq c_1M - \epsilon.$$

Therefore, there exist  $c, T > 0$  such that  $x_1(t) \geq cM$  for all  $t > T$ .  $\square$

Because  $x_1(t) \geq cM$ , provided  $cM > N$  we have  $x_1(t) \geq cM > N \geq y_1(t)$ , i.e., the system will finally become linear. In the following we will show how to apply our method to more general cases.

## 7.2. Convergence for two component types and one synchronisation (II): general case

This section deals with such an arbitrary PEPA model which has two component types and one synchronisation. The local action rates of the shared activity are not assumed to be the same. The main result of this section is a convergence theorem: as long as the population of one component type is sufficiently larger than the population of the other, then the solution of the derived ODEs converges as time tends to infinity.

### 7.2.1. Features of coefficient matrix

We assume the component types to be  $X$  and  $Y$ . The component type  $X$  is assumed to have local derivatives  $X_1, X_2, \dots, X_m$ , while  $Y$  has local derivatives  $Y_1, Y_2, \dots, Y_n$ . We use  $x_i(t)$  to denote the population of  $X$  in  $X_i$  ( $i = 1, \dots, m$ ) at time  $t$ . Similarly,  $y_j(t)$  denotes the population of  $Y$  in  $Y_j$  ( $j = 1, \dots, n$ ) at time  $t$ . Without loss of generality, we assume the synchronisation is associated with the local derivatives  $X_1$  and  $Y_1$ , i.e. the nonlinear term in the derived ODEs is  $\min\{rx_1(t), sy_1(t)\}$  where  $r$  and  $s$  are some constants. In fact, if the synchronisation is associated with  $X_i$  and  $Y_j$ , by appropriately permuting their suffixes, i.e.  $i \rightarrow 1, i+1 \rightarrow 2, \dots, i-1 \rightarrow m, j \rightarrow 1, j+1 \rightarrow 2, \dots, j-1 \rightarrow n$ , the synchronisation will be associated with  $X_1$  and  $Y_1$ . According to the mapping semantics presented previously, the derived ODEs from this class of PEPA model are

$$\frac{d\mathbf{x}}{dt} = \sum_l lf(\mathbf{x}, l) \quad (87)$$

where  $\mathbf{x} = (x_1(t), \dots, x_m(t), y_1(t), \dots, y_n(t))^T$ . For convenience, we denote

$$X(t) = (x_1(t), x_2(t), \dots, x_m(t))^T,$$

$$Y(t) = (y_1(t), y_2(t), \dots, y_n(t))^T.$$

In (87) all terms are linear except for those containing “ $\min\{rx_1(t), sy_1(t)\}$ ”. Notice

$$\min\{rx_1(t), sy_1(t)\} = I_{\{rx_1(t) \leq sy_1(t)\}} rx_1(t) + I_{\{rx_1(t) > sy_1(t)\}} sy_1(t).$$

When  $rx_1(t) \leq sy_1(t)$ , which is indicated by  $I_{\{rx_1(t) \leq sy_1(t)\}} = 1$  and  $I_{\{rx_1(t) > sy_1(t)\}} = 0$ , we can replace  $\min\{rx_1(t), sy_1(t)\}$  by  $rx_1(t)$  in (87). Then (87) becomes linear since all nonlinear terms are replaced by linear terms  $rx_1(t)$ , so the ODEs have the following form,

$$\begin{pmatrix} \frac{dX}{dt} \\ \frac{dY}{dt} \end{pmatrix} = Q_1 \begin{pmatrix} X \\ Y \end{pmatrix}, \quad (88)$$

where  $Q_1$  is a coefficient matrix. Similarly, if  $rx_1(t) > sy_1(t)$ ,  $\min\{rx_1(t), sy_1(t)\}$  can be replaced by  $sy_1(t)$  in (87). Then (87) can become

$$\begin{pmatrix} \frac{dX}{dt} \\ \frac{dY}{dt} \end{pmatrix} = Q_2 \begin{pmatrix} X \\ Y \end{pmatrix}, \quad (89)$$

where  $Q_2$  is another coefficient matrix corresponding to the case of  $rx_1(t) > sy_1(t)$ .

In short, the derived ODEs (87) are just the following

$$\begin{pmatrix} \frac{dX}{dt} \\ \frac{dY}{dt} \end{pmatrix} = I_{\{rx_1 \leq sy_1\}} Q_1 \begin{pmatrix} X \\ Y \end{pmatrix} + I_{\{rx_1 > sy_1\}} Q_2 \begin{pmatrix} X \\ Y \end{pmatrix}. \quad (90)$$

The case discussed in the previous section is a special case of this kind of form. If the conditions  $rx_1(t) \leq sy_1(t)$  and  $rx_1(t) > sy_1(t)$  occur alternately, then the matrices  $Q_1$  and  $Q_2$  will correspondingly alternately dominate the system, as Figure 7 illustrates.

Similar to the cases discussed in the previous two sections, the convergence problem of (90) can be divided into two subproblems, i.e. to examine whether the following two properties hold:

1. There exists a time  $T$ , such that either  $x_1 \leq y_1, \forall t > T$  or  $x_1 \geq y_1, \forall t > T$ .
2. The eigenvalues of  $Q_1$  and  $Q_2$  other than zeros have negative real parts.

The first item can guarantee (90) to eventually have a constant linear form, while the second item ensures the convergence of the bounded solution of the linear ODEs. If the answers to these two problems are both positive, then the convergence of the solution of (90) will hold. The study of these two problems are discussed in the next two subsections. In the remainder of this subsection, we first investigate the structure property of the coefficient matrices  $Q_1$  and  $Q_2$  in (90).

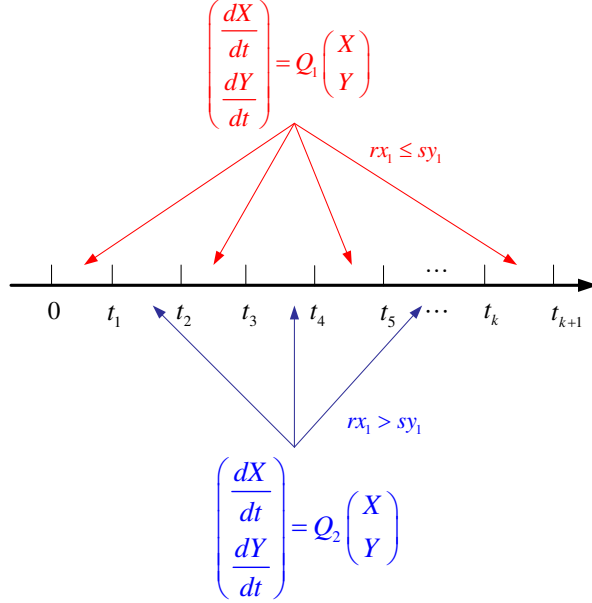


Figure 7: Illustration of derived ODEs with component types and one synchronisation

The structure of the coefficient matrices  $Q_1$  and  $Q_2$  is determined by the following two propositions, which indicate that they are either block lower-triangular or block upper-triangular.

**Proposition 14.**  $Q_1$  in (90) can be written as

$$Q_1 = \begin{pmatrix} \hat{Q}_1 & 0 \\ W & V_{n \times n} \end{pmatrix}_{(m+n) \times (m+n)}, \quad (91)$$

where  $\hat{Q}_1^T$  is an infinitesimal generator matrix with the dimension  $m \times m$ , and

$$W_{n \times m} = \begin{pmatrix} w_{11} & 0 & \cdots & 0 \\ w_{21} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ w_{n1} & 0 & \cdots & 0 \end{pmatrix}, \quad (92)$$

where  $w_{11} < 0$ ,  $w_{j1} (j = 2, \dots, n) \geq 0$  and  $\sum_{j=1}^n w_{j1} = 0$ . Here  $V$  and  $W$  satisfy that if we let

$$P = (W_1 + V_1, V_2, \dots, V_n), \quad (93)$$

i.e.  $P$ 's first column is the sum of  $V$ 's first column and  $W$ 's first column, while  $P$ 's other columns are the same to  $V$ 's other columns, then  $P^T$  is also an infinitesimal generator matrix.

*Proof.* Let

$$Q_1 = \begin{pmatrix} \hat{Q}_1 & U \\ W & V \end{pmatrix}, \quad (94)$$

where  $\hat{Q}_1$  and  $V$  are  $m \times m$  and  $n \times n$  matrices respectively. Suppose  $rx_1(t) \leq sy_1(t)$ , then

$$\begin{pmatrix} \frac{dX}{dt} \\ \frac{dY}{dt} \end{pmatrix} = Q_1 \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \hat{Q}_1 & U \\ W & V \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}. \quad (95)$$

So we have

$$\frac{dX}{dt} = \hat{Q}_1 X + UY. \quad (96)$$

The condition  $rx_1(t) \leq sy_1(t)$  implies that all nonlinear terms  $\min\{rx_1(t), sy_1(t)\}$  can be replaced by  $x_1(t)$ . This means that the behaviour of the component type  $X$  in (95) and (96) is independent of the component type  $Y$ . Thus in (95)  $U$  must be a zero matrix, i.e.

$$Q_1 = \begin{pmatrix} \hat{Q}_1 & 0 \\ W & V \end{pmatrix}.$$

Moreover, (96) becomes

$$\frac{dX}{dt} = \hat{Q}_1 X, \quad (97)$$

that is, there is no synchronisation in the ODEs corresponding to the component type  $X$  given  $rx_1(t) \leq sy_1(t)$ . Then by Proposition 4,  $\hat{Q}_1^T$  is an infinitesimal generator.

According to (95),

$$\begin{aligned} \frac{dY}{dt} &= WX + VY \\ &= (W_1, W_2, \dots, W_m)(x_1, x_2, \dots, x_m)^T + VY \\ &= x_1 W_1 + VY + \sum_{i=2}^m x_i W_i, \end{aligned} \quad (98)$$

where  $W = (W_1, W_2, \dots, W_m)$ . Notice that the component type  $Y$  is synchronised with the component type  $X$  only through the term  $\min\{rx_1(t), sy_1(t)\} = x_1(t)$ . In other words, in (98)  $Y$  is directly dependent on only  $x_1$  other than  $x_i$  ( $i \geq 2$ ). This implies  $W_1 \neq 0$  while  $W_i = 0$  ( $i = 2, 3, \dots, m$ ). Therefore,

$$\frac{dY}{dt} = x_1 W_1 + VY = x_1 W_1 + \sum_{j=1}^n y_j V_j, \quad (99)$$

where  $V_j$  ( $j = 1, 2, \dots, n$ ) are the columns of  $V$ . Denote  $W_1 = (w_{11}, w_{21}, \dots, w_{n1})^T$ . Notice that  $Y_1$  is a pre local derivative of the shared activity, and  $x_1 w_{11}$  represents the exit rates of the shared activity from  $Y_1$ . Therefore,  $w_{11} < 0$ . Moreover,  $x_1 w_{j1}$  ( $j = 2, \dots, n$ ) are the synchronised entry rates for the local derivatives  $Y_j$  ( $j = 2, \dots, n$ ) respectively, so  $w_{j1} \geq 0$  ( $j = 2, \dots, n$ ). By the conservation law, the total synchronised exit rates are equal to the total synchronised entry rates, i.e.  $x_1 \sum_{j=1}^n w_{j1} = 0$  or  $\sum_{j=1}^n w_{j1} = 0$ .

We have known that  $x_1$  in (99) derives from the synchronised term  $\min\{rx_1, sy_1\}$ . If the effect of the synchronisation on the behaviour of  $Y$  is removed, i.e. recover  $y_1$  by replacing  $x_1$ , then (99) will become

$$\frac{dY}{dt} = y_1 W_1 + VY = y_1 W_1 + \sum_{j=1}^n y_j V_j = PY, \quad (100)$$

where  $P = (W_1 + V_1, V_2, \dots, V_n)$ . Since there is no synchronisation contained in the subsystem of the component type  $Y$ , according to Proposition 4,  $P^T$  is the infinitesimal generator.  $\square$

Similarly, we can prove

**Proposition 15.**  $Q_2$  in (90) can be written as

$$Q_2 = \begin{pmatrix} E_{m \times m} & F \\ 0 & \hat{Q}_2 \end{pmatrix}_{(m+n) \times (m+n)}, \quad (101)$$

where  $\hat{Q}_2^T$  is an infinitesimal generator matrix with the dimension  $n \times n$ , and

$$F_{m \times n} = \begin{pmatrix} f_{11} & 0 & \cdots & 0 \\ f_{21} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ f_{m1} & 0 & \cdots & 0 \end{pmatrix}, \quad (102)$$



where  $f_{11} < 0$ ,  $f_{j1}(j = 2, \dots, m) \geq 0$  and  $\sum_{j=1}^m f_{j1} = 0$ . Here  $F$  and  $E$  satisfy that if we let

$$R = (F_1 + E_1, E_2, \dots, E_n), \quad (103)$$

then  $R^T$  is also an infinitesimal generator matrix.

### 7.2.2. Eigenvalues of coefficient matrix

In this subsection, we will determine the eigenvalue property of  $Q_1$  and  $Q_2$ . First, the Perron-Frobenius theorem gives an estimation of eigenvalues for nonnegative matrices.

**Theorem 6. (*Perron-Frobenius*).** Let  $A = (a_{ij})$  be a real  $n \times n$  matrix with nonnegative entries  $a_{ij} \geq 0$ . Then the following statements hold:

1. There is a real eigenvalue  $r$  of  $A$  such that any other eigenvalue  $\lambda$  satisfies  $|\lambda| \leq r$ .
2.  $r$  satisfies  $\min_i \sum_j a_{ij} \leq r \leq \max_i \sum_j a_{ij}$ .

**Remark 5.** We should point out that in the second property, exchanging  $i$  and  $j$  in  $a_{ij}$  in the formula, we still have  $\min_i \sum_j a_{ji} \leq r \leq \max_i \sum_j a_{ji}$ . In fact,  $A^T$  is also a real matrix with non-negative entries. Since  $A^T$  and  $A$  share the same eigenvalues, so  $r$  is one of the eigenvalues of  $A^T$ , such that any other eigenvalue  $\lambda$  of  $A^T$  satisfies  $|\lambda| \leq r$ . Notice  $(A^T)_{ij} = A_{ji}$ , By applying the Perron-Frobenius theorem to  $A^T$ , we have

$$\min_i \sum_j a_{ji} \leq r \leq \max_i \sum_j a_{ji}.$$

We cannot directly apply this theorem to our coefficient matrices  $Q_1$  and  $Q_2$ , since both of them have negative elements, not only on the diagonal but also in other entries. However, we use some well-known techniques in linear algebra, i.e. the following Lemma 10 and 11 (which can be easily found in linear algebra textbooks), to cope with this problem, and thus derive estimates of their eigenvalues.

**Lemma 10.** If  $E_{m \times m}$  and  $F_{n \times n}$  have eigenvalues  $\lambda_i$  ( $i = 1, 2, \dots, m$ ) and  $\delta_j$  ( $j = 1, 2, \dots, n$ ) respectively, then each of

$$H_1 = \begin{pmatrix} E & 0 \\ G & F \end{pmatrix}, \quad H_2 = \begin{pmatrix} E & G \\ 0 & F \end{pmatrix}$$

has eigenvalues  $\lambda_i$  ( $i = 1, 2, \dots, m$ ) and  $\delta_j$  ( $j = 1, 2, \dots, n$ ).

**Lemma 11.** *If  $\lambda$  is an eigenvalue of  $V$ , then  $\lambda + r$  is an eigenvalue of  $V + rI$ , where  $r$  is a scalar.*

**Theorem 7.** *The eigenvalues of both  $Q_1$  and  $Q_2$  are either zeros or have negative real parts.*

*Proof.* We only give the proof for  $Q_1$ 's case. By Proposition 14,

$$Q_1 = \begin{pmatrix} \hat{Q}_1 & 0 \\ W & V_{n \times n} \end{pmatrix}_{(m+n) \times (m+n)}. \quad (104)$$

According to Lemma 10, if all eigenvalues of  $\hat{Q}_1$  and  $V$  are determined, then the eigenvalues of  $Q_1$  will be determined. Let us consider  $V$  first.

Notice that only diagonal elements of  $V$  are possibly negative (which can be deduced from Proposition 14). Let  $r = \sup_i |V_{ii}| > 0$ , then all the entries of  $V + rI$  are nonnegative. Let  $\lambda$  be an arbitrary eigenvalue of  $V$ , then by Lemma 11,  $\lambda + r$  is an eigenvalue of  $V + rI$ .

Notice the sum of the elements of any column of  $V$  is zero (because the sum of entry rates equals to the sum of exit rates), so  $V$  has a zero eigenvalue with the corresponding eigenvector  $\mathbf{1}$ , i.e.  $V\mathbf{1} = \mathbf{0}$ . Thus  $r = 0 + r$  is an eigenvalue of  $V + rI$ . Moreover,

$$\min_i \sum_j (V + rI)_{ji} = r = \max_i \sum_j (V + rI)_{ji}.$$

Applying the Perron-Frobenius theorem (Theorem 6) and Remark 5 to  $V + rI$ , so

$$|\lambda + r| \leq r. \quad (105)$$

Let  $\lambda = a + bi$ , then (105) implies that  $a \leq 0$ , and if  $a = 0$  then  $b = 0$ . In other words,  $V$ 's eigenvalues are either zeros or have negative real parts.

Similarly,  $\hat{Q}_1$ 's eigenvalues other than zeros have negative real parts. By Lemma 10, the eigenvalues of  $Q_1$  other than zeros have negative real parts. The proof is complete.  $\square$

### 7.2.3. Convergence theorem

Now we deal with another subproblem: whether or not after a long time, we always have  $rx_1 \geq sy_1$  (or  $rx_1 < sy_1$ ). If the population of  $X$  is significantly larger than the population of  $Y$ , intuitively, there will finally be a greater number of  $X$  in the local derivative  $X_1$ , than the number of  $Y$  in  $Y_1$ . This will lead to  $rx_1 > sy_1$ .

**Lemma 12.** *Under the assumptions in Section 7.1.1, for the following ODEs*

$$\begin{pmatrix} \frac{dX}{dt} \\ \frac{dY}{dt} \end{pmatrix} = I_{\{rx_1 \leq sy_1\}} Q_1 \begin{pmatrix} X \\ Y \end{pmatrix} + I_{\{rx_1 > sy_1\}} Q_2 \begin{pmatrix} X \\ Y \end{pmatrix},$$

*there exists  $c_1 > 0, c_2 > 0, T > 0$ , such that  $x_1(t) \geq c_1 M$ ,  $y_1(t) \geq c_2 N$  for any  $t \geq T$ , where  $c_1$  and  $c_2$  are independent of  $M$  and  $N$ .*

*Proof.* The proof is essentially the same as the proof of Lemma 9. We only give the sketch of the proof for  $x_1(t) \geq c_1 M$ . By introducing new two functions  $\alpha(t)$  and  $\beta(t)$ ,

$$\alpha(t) = \begin{cases} \frac{\min\{rx_1(t), sy_1(t)\}}{rx_1(t)}, & x_1(t) \neq 0, \\ 1, & x_1(t) = 0, \end{cases}$$

$$\beta(t) = \frac{1}{t} \int_0^t \alpha(s) ds,$$

the nonlinear term  $\min\{rx_1(t), sy_1(t)\}$  equals  $r\alpha(t)x_1(t)$ , and thus the ODEs associated with the subsystem  $X$  can be written as

$$\frac{dX}{dt} = A(t)X \tag{106}$$

where  $A(t)$  is related to  $\alpha(t)$ . The solution of (106) is  $X(t) = e^{tB(t)}X(0)$ , where  $B(t)$  is defined by  $B(t) = \frac{1}{t} \int_0^t A(s) ds$ , and thus  $B(t)$  is related to  $\beta(t)$ .

Notice that according to Theorem 7 and its proof, the eigenvalues of  $B(t)$  other than zeros have negative real parts for any  $t > 0$ . By a similar proof to Corollary 1, we have

$$\Lambda = \inf_{t \geq 0} \{-\Re(\lambda) \mid \lambda \text{ is } B(t)\text{'s non-zero eigenvalue}\} > 0. \tag{107}$$

This fact will lead to the conclusion that  $X(t)$  can be approximated by  $\hat{X}(t)$ , where  $\hat{X}(t)$  is constructed similarly to the one in Proposition 12. Because a general  $B(t)$  considered here may not be diagonalisable, so the construction of  $\hat{X}(t)$  is a little bit more complicated. We detail the construction as well as the proof of the following result in Appendix D:

$$\lim_{t \rightarrow \infty} \|X(t) - \hat{X}(t)\| = 0.$$

Then, by similar arguments to Proposition 13 and Corollary 2, we can prove that  $\inf_{t>T} \hat{x}_1(t) \geq cM$ , where  $\hat{x}_1(t)$  is the first entry of  $\hat{X}(t)$ . Then, by a similar proof to the proof of Lemma 9, we can conclude that there exists a number  $c_1$  such that  $x_1(t) \geq c_1M$  after a time  $T$ .  $\square$

**Lemma 13.** *Under the assumptions of Lemma 12, if  $M > K_1N$  or  $N > K_2M$ , where constants  $K_1 > 0$  and  $K_2 > 0$  are sufficiently large, then there exists  $T$  such that  $rx_1(t) \geq sy_1(t)$ ,  $\forall t \geq T$  or  $rx_1(t) \leq sy_1(t)$ ,  $\forall t \geq T$  respectively.*

*Proof.* By the boundedness of solutions,  $0 \leq x_1(t) \leq M$  and  $0 \leq y_1(t) \leq N$  for any  $t$ . Suppose  $M > K_1N$ . By Lemma 12, there exists  $c, T > 0$ ,  $x_1(t) \geq cM$ ,  $\forall t \geq T$ . Since  $K_1$  is assumed to be large enough such that  $K_1 \geq \frac{s}{rc}$ , then  $rcM > sN$ . So for any  $t > T$ , we have

$$rx_1(t) \geq rcM \geq sN \geq sy_1(t).$$

If  $N > K_2M$ , the proof is similar and omitted here.  $\square$

Now we state our convergence theorem.

**Theorem 8.** *If  $M > K_1N$  or  $N > K_2M$ , where constants  $K_1, K_2 > 0$  are sufficiently large, then the solution of the derived ODEs (90), i.e.*

$$\begin{pmatrix} \frac{dX}{dt} \\ \frac{dY}{dt} \end{pmatrix} = I_{\{rx_1 \leq sy_1\}} Q_1 \begin{pmatrix} X \\ Y \end{pmatrix} + I_{\{rx_1 > sy_1\}} Q_2 \begin{pmatrix} X \\ Y \end{pmatrix},$$

*converges to a finite limit as time goes to infinity.*

*Proof.* Suppose  $M > K_1N$ , then by Lemma 13, there exists a time  $T > 0$ , such that after time  $T$ ,  $rx_1(t) \geq sy_1(t)$ , so (90) becomes

$$\begin{pmatrix} \frac{dX}{dt} \\ \frac{dY}{dt} \end{pmatrix} = Q_2 \begin{pmatrix} X \\ Y \end{pmatrix}. \quad (108)$$

Since  $Q_2$ 's eigenvalues other than zeros have strict negative real parts according to Theorem 7, and the solution of the above equation is bounded, then by Corollary 3, the solution converges to a finite limit as time goes to infinity.  $\square$

**Remark 6.** *We should point out that for a general PEPA model with two component types and one synchronisation, the limit of the solution of the derived ODEs is determined by the populations of these two component types rather than the particular starting state. That is to say, whatever the initial state is, as long as the total populations of the components are the same, the limit that the solution converges to will always be the same. We do not plan to discuss this topic in detail in this paper.*

## 8. Related work

As we have mentioned in the introduction section, the fluid approximation approach was first proposed by Hillston in [21] to deal with the state space explosion problem in the context of PEPA. An interpretation as well as a justification of this approximation approach has been demonstrated by Hayden in his dissertation [33]. In [33, 34], generation of similar systems of coupled ODEs for higher-order moments such as variance has been addressed. Additionally, the dissertation [33] discusses how to derive stochastic differential equations from PEPA models.

More recently, some extensions of the previous mapping from PEPA to ODEs have been presented by Bradley *et al.* in [23]. In particular, passive rates are introduced into the fluid approximation. In the recent paper [35], different existing styles of passive cooperation in fluid models are compared and intensively discussed. Moreover, a new passive fluid semantics for passive cooperation, which can be viewed as approximating the first moments of the component counting processes, has been provided, with a theoretical justification. The paper [36] considers the application of this fluid approximation approach with modifications in the context of epidemiology. In this paper, the notions of side and self-loops are added to the activity matrix, and the rates are calculated differently, for the purpose of deriving from PEPA models the most commonly used ODEs in the context of epidemiology. In [37] and [38] by Tribastone, a new operational semantics is proposed to give a compact symbolic representation of PEPA models. This semantics extends the application scope of the fluid approximation of PEPA by incorporating all the operators of the language and removing earlier assumptions on the syntactical structure of the models amenable to this analysis.

The fluid approximation approach has also been applied to timed Petri nets to deal with the state space explosion problem [39, 40]. The comparison between the fluid approximation of PEPA models and timed continuous Petri

nets has been demonstrated by Galpin in [41]. This paper has established links between two continuous approaches to modelling the performance of systems. In the paper, a translation from PEPA models to continuous Petri nets and *vice versa* has been presented. In addition, it has been shown that the continuous approximation using PEPA has infinite server semantics. The fluid approximation approach has also been used by Thomas to derive asymptotic solutions for a class of closed queueing networks [42]. In this paper, an analytical solution to a class of models, specified using PEPA, is derived through the ODE approach. It is shown that “this solution is identical to that used for many years as an asymptotic solution to the mean value analysis of closed queueing networks”.

Moreover, the relationship between the fluid approximation and the underlying CTMCs for a special PEPA model has been revealed by Geisweiller *et al.* in [25]: the ODEs derived from the PEPA description are the limits of the sequence of underlying CTMCs. It has been shown in [28] by Gilmore that for some special examples the equilibrium points of the ODEs derived from PEPA models coincide the steady-state probability distributions of the CTMCs underlying the nonsynchronised PEPA models.

In addition, there are several papers which discuss how to derive response time from the fluid approximation of PEPA models. In [43], by constructing an absorption operator for the PEPA language, Bradley *et al.* allow general PEPA models to be analysed for fluid-generated response times. Clark *et al.* demonstrate in [44] how to derive expected passage response times using Little’s law based on averaged populations of entities in an equilibrium state. This technique has been generalised into one for obtaining a full response-time profile computing the probability of a component observing the completion of a response at a given time after the initiation of the request, see [45]. Moreover, an error in the passage specification in the approach taken in [43] has been uncovered and rectified in [45] by Clark. The ODE method associated with the PEPA language has demonstrated successful application in the performance analysis of large scale systems [46, 47, 48, 49, 50].

## 9. Conclusions

In this paper, we have demonstrated how to derive the fluid approximation from a general PEPA model via the numerical representation of the PEPA language, which extended the current mapping semantics of fluid approximations. The fundamental properties of the fluid approximation such

as the existence and uniqueness, boundedness and nonnegativeness of the solutions of the derived ODEs have been established. Moreover, the convergence of the solution under a particular condition for general models has been verified. This particular condition relates some famous constants of Markov chains such as the spectral gap and the Log-Sobolev constant. For the models without synchronisations or with one synchronisation and two component types, we have determined the convergence. Furthermore, the consistency between the fluid approximation and the CTMC in the context of PEPA has been revealed. If a model has no synchronisations, then the derived ODEs are just the probability distribution evolution equations of the underlying CTMC except for a scaling factor. For any general PEPA model, the ODEs can be taken as the corresponding density dependent CTMC with the concentration level infinity. In addition, the coefficient matrices of some derived ODEs were studied: their eigenvalues are either zeros or have negative real parts. The structural property of invariance has been shown to play an important role in the proof of convergence for some PEPA models. Due to the limitation of pages, we have not shown how to derive performance measures from the fluid approximation, and have not given the numerical comparison between the fluid approximation and the CTMC in terms of performance measures. For more details, please refer to [22].

In addition to the established results, this paper has also demonstrated comprehensive techniques and methods to investigate the PEPA language: not only both theoretical and experimental, probabilistic and analytic, but also syntactic and numerical (investigating the models based on the numerical representation of PEPA), qualitative and quantitative (exploiting the structural property to verify the convergence). These results and techniques are expected to have more applications in performance modelling and evaluation of large scale systems.

## Acknowledgements

Partial work of the first author was carried out in LFCS, School of Informatics and IDCom, School of Engineering, The University of Edinburgh, when he was a PhD student and supported by the Mobile VCE ([www.mobilevce.com](http://www.mobilevce.com)) Programme 4.

- [1] J. Hillston, A Compositional Approach to Performance Modelling (PhD Thesis), Cambridge University Press, 1996.

- [2] N. Götz, U. Herzog, M. Rettelsbach, TIPP– a language for timed processes and performance evaluation, Tech. rep., Tech. Rep.4/92, IMMD7, University of Erlangen-Nürnberg, Germany (Nov. 1992).
- [3] M. Bernardo, R. Gorrieri, A tutorial on EMPA: A theory of concurrent processes with nondeterminism, priorities, probabilities and time, *Theoretical Computer Science* 202 (1998) 1–54.
- [4] H. Hermanns, Interactive Markov chains, Ph.D. thesis, Universität Erlangen-Nürnberg, Germany (1998).
- [5] M. Sereno, Towards a product form solution for stochastic process algebras, *The Computer Journal* 38 (7) (1995) 622–632.
- [6] J. Hillston, N. Thomas, Product form solution for a class of PEPA models, *Performance Evaluation* 35 (3–4) (1999) 171–192.
- [7] G. Clark, J. Hillston, Product form solution for an insensitive stochastic process algebra structure, *Performance Evaluation* 50 (2–3) (2002) 129–151.
- [8] P. G. Harrison, Turning back time in Markovian process algebra, *Theor. Comput. Sci.* 290 (3) (2003) 1947–1986.
- [9] V. Mertsiotakis, Approximate analysis methods for stochastic process algebras, Ph.D. thesis, Universität Erlangen-Nürnberg, Erlangen (1998).
- [10] H. C. Bohnenkamp, B. R. Haverkort, Semi-numerical solution of stochastic process algebra models, in: ARTS '99: Proceedings of the 5th International AMAST Workshop on Formal Methods for Real-Time and Probabilistic Systems, Springer-Verlag, London, UK, 1999, pp. 228–243.
- [11] J. Hillston, L. Kloul, An efficient Kronecker representation for PEPA models, in: PAPM-PROBMIV '01: Proceedings of the Joint International Workshop on Process Algebra and Probabilistic Methods, Performance Modeling and Verification, Springer-Verlag, London, UK, 2001, pp. 120–135.
- [12] J. Hillston, L. Kloul, Formal techniques for performance analysis: blending SAN and PEPA, *Formal Aspects of Computing* 19 (2007) 3–33.



- [13] B. Buchholz, Compositional analysis of a Markovian process algebra, in: M. Rettelsbach, U. Herzog (Eds.), Proc. of 2nd Process Algebra and Performance Modelling Workshop, 1994.
- [14] M. Rettelsbach, M. Siegle, Compositional minimal semantics for the stochastic process algebra TIPP, in: Proc. of 2nd Process Algebra and Performance Modelling Workshop, 1994, p. pages.
- [15] M. J. A. Smith, Abstraction and model checking in the Eclipse PEPA plug-in, in: PASTA '09: Proceedings of the 8th Workshop on Process Algebra and Stochastically Timed Activities, 2009.
- [16] M. J. A. Smith, Compositional Abstraction of PEPA Models, [http://lanther.co.uk/papers/PEPA\\_abstraction.pdf](http://lanther.co.uk/papers/PEPA_abstraction.pdf) (2009).
- [17] C. Priami, A. Regev, E. Shapiro, W. Silverman, Application of a stochastic name-passing calculus to representation and simulation of molecular processes, *Inf. Process. Lett.* 80 (1) (2001) 25–31.
- [18] C. Kuttler, J. Niehren, Gene regulation in the  $\pi$ -calculus: Simulating cooperativity at the lambda switch, Vol. 4230 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, 2006, pp. 24–55.
- [19] M. Calder, A. Duguid, S. Gilmore, J. Hillston, Stronger computational modelling of signalling pathways using both continuous and discrete-state methods, in: *Computational Methods in Systems Biology*, Vol. 4210 of LNCS, Springer-Verlag, 2006, pp. 63–77.
- [20] M. Calder, S. Gilmore, J. Hillston, Modelling the influence of RKIP on the ERK signalling pathway using the stochastic process algebra PEPA, in: *Transactions on Computational Systems Biology VII*, no. 4230 in LNCS, Springer, 2006.
- [21] J. Hillston, Fluid flow approximation of PEPA models, in: *International Conference on the Quantitative Evaluation of Systems (QEST'05)*, IEEE Computer Society, 2005.
- [22] J. Ding, Structural and fluid analysis of large scale PEPA models — with applications to content adaptation systems, Ph.D. thesis, The University of Edinburgh (2010).

- [23] J. T. Bradley, S. T. Gilmore, J. Hillston, Analysing distributed internet worm attacks using continuous state-space approximation of process algebra models, *Journal of Computer and System Sciences* 74 (6) (2008) 1013–1032.
- [24] S. Gilmore, J. Hillston, M. Ribaud, An efficient algorithm for aggregating PEPA models, *IEEE Trans. Softw. Eng.* 27 (5) (2001) 449–464.
- [25] N. Geisweiller, J. Hillston, M. Stenico, Relating continuous and discrete PEPA models of signalling pathways, *Theoretical Computer Science* 404 (1-2) (2008) 97–111.
- [26] G. Bolch, S. Greiner, H. d. Meer, K. S. Trivedi, *Queueing Networks and Markov Chains: Modelling and Performance Evaluation with Computer Science Application*, John Wiley & Sons, INC., 1998.
- [27] J. H. Hubbard, B. H. West, *Differential Equations: A Dynamical Systems Approach (Higher-Dimensional Systems)*, no. 18 in *Texts in Applied Mathematics*, Springer, 1990.
- [28] S. Gilmore, Continuous-time and continuous-space process algebra, in: *Process Algebra and Stochastically Timed Activities (PASTA’05)*, 2005.
- [29] T. G. Kurtz, Solutions of ordinary differential equations as limits of pure jump Markov processes, *Journal of Applied Probability* 7 (1) (1970) 49–58.
- [30] S. N. Ethier, T. G. Kurtz, *Markov Processes: Characterization and Convergence*, John Wiley & Sons, Inc., 1986.
- [31] Z. Brzeźniak, T. Zastawniak, *Basic Stochastic Processes*, Springer Undergraduate Mathematics Series, Springer, 1999.
- [32] L. Saloff-Coste, Lectures on finite Markov chains, in: *Lecture Notes in Mathematics*, Vol. 1665, Springer, Berlin, 1997.
- [33] R. Hayden, Addressing the state space explosion problem for PEPA models through fluid-flow approximation, <http://pubs.doc.ic.ac.uk/fluid-spa-modelling/> (July 2007).

- [34] R. A. Hayden, J. T. Bradley, ODE-based general moment approximations for PEPA, in: Proceedings of 7th Workshop on Process Algebra and Stochastically Timed Activities (PASTA'08), 2008.
- [35] R. Hayden, J. T. Bradley, Evaluating fluid semantics for passive stochastic process algebra cooperation, *Performance Evaluation* 67 (4) (2010) 260–284.
- [36] S. Benkirane, J. Hillston, C. McCaig, R. Norman, C. Shankland, Improved continuous approximation of PEPA models through epidemiological examples, *Electron. Notes Theor. Comput. Sci.* 229 (1) (2009) 59–74.
- [37] M. Tribastone, Differential analysis of PEPA models, in: Proceedings of 8th Workshop on Process Algebra and Stochastically Timed Activities (PASTA'09), 2009.
- [38] M. Tribastone, S. Gilmore, J. Hillston, Scalable Differential Analysis of Process Algebra Models, to appear in *IEEE Transactions on Software Engineering* (2010).
- [39] M. Silva, L. Recalde, Continuization of timed Petri nets: From performance evaluation to observation and control., in: Lecture Notes in Computer Science: Applications and Theory of Petri Nets 2005: 26th International Conference, ICATPN 2005, Miami, USA, June 20-25, 2005. / Gianfranco Ciardo, Philippe Darondeau (Eds.), Vol. 3536, Springer Verlag, 2005, pp. 26–47.
- [40] C. Mahulea, L. Recalde, M. Silva, On performance monotonicity and basic servers semantics of continuous Petri nets, Ann Arbor, USA.
- [41] V. Galpin, Continuous approximation of PEPA models and Petri nets, in: Proceedings of the European Simulation and Modelling Conference (ESM 2008), Le Havre, France, 2008, pp. 492–499.
- [42] N. Thomas, Using ODEs from PEPA models to derive asymptotic solutions for a class of closed queueing networks, in: PASTA '09: Proceedings of the 8th Workshop on Process Algebra and Stochastically Timed Activities, 2009.

- [43] J. T. Bradley, R. Hayden, W. J. Knottenbelt, T. Suto, Extracting Response Times from Fluid Analysis of Performance Models, in: SIPEW'08, SPEC International Performance Evaluation Workshop, Darmstadt, 27-28 June 2008, Vol. 5119 of Lecture Notes in Computer Science, 2008, pp. 29–43.
- [44] A. Clark, A. Duguid, S. Gilmore, J. Hillston, Espresso, a little coffee, in: PASTA '08: Proceedings of the 7th Workshop on Process Algebra and Stochastically Timed Activities, 2008.
- [45] A. Clark, Response-time profiles for PEPA models compiled to ODEs, in: PASTA '09: Proceedings of the 8th Workshop on Process Algebra and Stochastically Timed Activities, 2009.
- [46] M. Harrison, M. Massink, Modelling interactive experience, function and performance in ubiquitous systems, in: Proc. of 4th International Workshop on Practical Applications of Stochastic Modelling (PASM'09), Imperial College, London, 2009.
- [47] Y. Zhao, N. Thomas, Approximate solution of a PEPA model of a key distribution centre, in: Performance Evaluation: Metrics, Models and Benchmarks, Vol. 5119 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2008, pp. 44–57.
- [48] N. Thomas, Y. Zhao, Fluid flow analysis of a model of a secure key distribution centre, in: Proceedings of 24th Annual UK Performance Engineering Workshop, Imperial College, London, 2008.
- [49] Y. Zhao, N. Thomas, Efficient solutions of a PEPA model of a key distribution centre, *Performance Evaluation* 67 (8) (2010) 740–756.
- [50] J. Ding, J. Hillston, D. Laurenson, Performance modelling of content adaptation for a personal distributed environment, *Wireless Personal Communications* 48 (2009) 93–112.
- [51] L. Perko, *Differential Equations and Dynamical Systems*, no. 7 in Texts in Applied Mathematics, Springer-Verlag, 1991.

## Appendix A.

**Theorem 9.** (*Kurtz theorem* [30], page 456). Let  $X_n$  be a family of density dependent CTMCs with the infinitesimal generators

$$q_{k,k+l}^{(n)} = nf(k/n, l),$$

where  $f(x, l)$  ( $x \in E \subset R^h$ ,  $l \in \mathbb{Z}^h$ ) is a continuous function,  $k$  is a numerical state vector and  $l$  is a transition vector.

Suppose  $X(t) \in E$  satisfies

$$\frac{dx}{dt} = F(x)$$

where  $F(x) = \sum_l lf(x, l)$ . Suppose that for each compact  $K \subset E$ ,

$$\sum_l \|l\| \sup_{x \in K} f(x, l) < \infty \quad (\text{A.1})$$

and there exists  $M_K > 0$  such that

$$\|F(x) - F(y)\| \leq M_K \|x - y\|, \quad x, y \in K. \quad (\text{A.2})$$

If  $\lim_{n \rightarrow \infty} \frac{X_n(0)}{n} = x_0$ , then for every  $t \geq 0$ ,

$$\lim_{n \rightarrow \infty} \sup_{s \leq t} \left\| \frac{X_n(s)}{n} - X(s) \right\| = 0 \quad a.s. \quad (\text{A.3})$$

The following lemma can be found in any good book on differential calculus.

**Lemma 14.** Let  $y(t)$  be a differentiable function defined for  $t \geq 0$ . Suppose  $a, b \in \mathbb{R}$ ,  $a \neq 0$ . If  $y(t)$  satisfies  $\frac{dy}{dt} \geq ay(t) + b$ ,  $t > 0$ , then

$$y(t) \geq e^{at} \left( y(0) + \frac{b}{a} \right) - \frac{b}{a}.$$

Similarly, if  $y(t)$  satisfies  $\frac{dy}{dt} \leq ay(t) + b$ ,  $t > 0$ , then

$$y(t) \leq e^{at} \left( y(0) + \frac{b}{a} \right) - \frac{b}{a}.$$

*Proof.* Let  $W(t) = y(t)e^{-at}$ , then  $\frac{dW}{dt} = e^{-at} \left( \frac{dy}{dt} - ay \right) \geq be^{-at}$ . Integrating on both sides, so  $W(t) - W(0) \geq b \int_0^t e^{-as} ds$ . Thus  $y(t)e^{-at} - y(0) \geq \frac{b}{a}(1 - e^{-at})$ . So  $y(t) \geq e^{at} \left( y(0) + \frac{b}{a} \right) - \frac{b}{a}$ . The second conclusion can be similarly proved.  $\square$

**Theorem 10.** (*Fundamental Inequality*, [27], page 14). If  $\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, t)$  is defined on a set  $U$  in  $\mathbb{R}^n \times \mathbb{R}$  with the Lipschitz condition

$$\|\mathbf{f}(\mathbf{x}_1, t) - \mathbf{f}(\mathbf{x}_2, t)\| < K\|\mathbf{x}_1 - \mathbf{x}_2\|$$

for all  $(\mathbf{x}_1, t)$  and  $(\mathbf{x}_2, t)$  on  $U$ , and if for  $\epsilon_i, \delta \in \mathbb{R}$ , and  $\mathbf{u}_1(t)$  and  $\mathbf{u}_2(t)$  are two continuous, piecewise differentiable functions on  $U$  into  $\mathbb{R}^n$  with

$$\left\| \frac{d\mathbf{u}_i(t)}{dt} - \mathbf{f}(\mathbf{u}_i(t), t) \right\| \leq \epsilon_i, \quad \text{and} \quad \|\mathbf{u}_1(t_0) - \mathbf{u}_2(t_0)\| \leq \delta,$$

then

$$\|\mathbf{u}_1(t) - \mathbf{u}_2(t)\| \leq \delta e^{K(t-t_0)} + \left( \frac{\epsilon_1 + \epsilon_2}{K} \right) (e^{K(t-t_0)} - 1).$$

## Appendix B.

The material presented here is extracted from [32]. Let  $(K, \pi)$  be a Markov chain on a finite set  $S$ , where  $K$  is a Markov kernel and  $\pi$  is the stationary probability distribution associated with  $K$ . For any real function  $f, g$  on  $S$ , define an inner product “ $\langle \cdot, \cdot \rangle$ ” as

$$\langle f, g \rangle = \sum_{x \in S} f(x)g(x)\pi(x).$$

Denote  $\|f\|_2 = \sqrt{\langle f, f \rangle}$ , and

$$l^2(\pi) = \{f : \|f\|_2 < \infty\}.$$

Then  $l^2(\pi)$  is a Hilbert space with the norm  $\|\cdot\|_2$ . We say that  $K^*$  is *adjoint* to  $K$  if

$$\langle Kf, g \rangle = \langle f, K^*g \rangle, \quad \forall f, g \in L^2(\pi).$$

It follows that

$$K^*(x, y) = \frac{\pi(y)}{\pi(x)} K(y, x).$$

If  $K = K^*$ , then  $K$  is called *self-adjoint*. If  $K$  is self-adjoint on  $l^2(\pi)$ , then  $(K, \pi)$  is *reversible*.

For a function in  $f \in l^2(\pi)$ , denote its mean and variance by  $\pi(f)$  and  $\text{Var}(f)$  respectively, that is

$$\pi(f) = \sum_{x \in S} f(x) \pi(x), \quad \text{Var}(f) = \pi((f - \pi(f))^2).$$

**Definition 8. (*Dirichlet form*).** The form

$$\mathcal{E}(f, g) = \langle (I - K)f, g \rangle$$

is called the *Dirichlet form* associated with  $H_t = e^{-t(I-K)}$ .

**Remark 7.** The Dirichlet form  $\mathcal{E}$  satisfies

$$\mathcal{E}(f, f) = \langle (I - K)f, f \rangle = \left\langle \left( I - \frac{K + K^*}{2} \right) f, f \right\rangle,$$

$$\mathcal{E}(f, f) = \frac{1}{2} \sum_{x, y} (f(x) - f(y))^2 K(x, y) \pi(x).$$

**Definition 9. (*Spectral gap*).** Let  $K$  be a Markov kernel with Dirichlet form  $\mathcal{E}$ . The spectral gap  $\lambda = \lambda(K)$  is defined by

$$\lambda = \min \left\{ \frac{\mathcal{E}(f, f)}{\text{Var}(f)} : \text{Var}(f) \neq 0 \right\}.$$

**Remark 8.** In general  $\lambda$  is the smallest non zero eigenvalue of  $I - \frac{K+K^*}{2}$ . If  $K$  is self-adjoint, then  $\lambda$  is the smallest non zero eigenvalue of  $I - K$ . Clearly, we also have

$$\lambda = \min \{ \mathcal{E}(f, f) : \|f\|_2 = 1, \pi(f) = 0 \}.$$

The definition of the logarithmic Sobolev (Log-Sobolev) constant  $\alpha$  is similar to that of the spectral gap  $\lambda$  where the variance has been replaced by

$$\mathcal{L}(f) = \sum_{x \in S} f(x)^2 \log \left( \frac{f(x)^2}{\|f\|_2^2} \right) \pi(x).$$

**Definition 10. (*Log-Sobolev constant*).** Let  $K$  be an irreducible Markov chain with stationary measure  $\pi$ . The logarithmic Sobolev constant  $\alpha = \alpha(K)$  is defined by

$$\alpha = \min \left\{ \frac{\mathcal{E}(f, f)}{\mathcal{L}(f)} : \mathcal{L}(f) \neq 0 \right\}.$$

**Lemma 15.** For any finite Markov chain  $K$  with stationary measure  $\pi$ , the Log-Sobolev constant  $\alpha$  and the spectral gap  $\lambda$  satisfy

$$\frac{1 - 2\pi(*)}{\log[1/\pi(*) - 1]} \lambda \leq \alpha \leq \frac{\lambda}{2},$$

where  $\pi(*) = \min_x \pi(x)$ .

## Appendix C.

In this appendix, we claim that all eigenvalues of  $Q_i$  ( $i = 1, 2, 3, 4$ ) appearing in (54) in Section 6.2.1 other than zeros have negative real parts. We do not worry about  $Q_1$  and  $Q_4$  since they are lower or upper block triangular matrices and the eigenvalues of this kind of matrices can be well estimated: all eigenvalues of  $Q_1$  and  $Q_4$  are either zeros or have negative real parts. All that we want to do here is to show that both  $Q_2$  and  $Q_3$  also have this property.

By symbolic calculation using Matlab,  $Q_3$ 's eigenvalues are

$$\lambda_{1,2,3} = 0(\text{three folds}), \lambda_4 = -c_4 - c_3,$$

$$\lambda_5 = -\frac{1}{2}(a_1 + a_2 + c_1 + c_2) + \frac{1}{2}\sqrt{(a_1 - a_2 + c_1 + c_2)^2 - 4a_1c_2},$$

$$\lambda_6 = -\frac{1}{2}(a_1 + a_2 + c_1 + c_2) - \frac{1}{2}\sqrt{(a_1 - a_2 + c_1 + c_2)^2 - 4a_1c_2}.$$

If  $(a_1 - a_2 + c_1 + c_2)^2 - 4a_1c_2 < 0$ , then the real parts of  $\lambda_5$  and  $\lambda_6$  are  $-\frac{1}{2}(a_1 + a_2 + c_1 + c_2)$ , which is negative. Otherwise,

$$(a_1 - a_2 + c_1 + c_2)^2 - 4a_1c_2 \geq 0.$$

In this case,

$$(a_1 - a_2 + c_1 + c_2)^2 - 4a_1c_2 \leq (a_1 - a_2 + c_1 + c_2)^2 < (a_1 + a_2 + c_1 + c_2)^2,$$



so

$$-\frac{1}{2}(a_1 + a_2 + c_1 + c_2) + \frac{1}{2}\sqrt{(a_1 - a_2 + c_1 + c_2)^2 - 4a_1c_2} < 0.$$

This means that  $\lambda_5$  and  $\lambda_6$  are both negative. Thus  $\lambda_i$  ( $i = 1, 2, \dots, 6$ ) are either 0 or have negative real parts.

Similarly,  $Q_2$ 's eigenvalues are  $\delta_{1,2,3} = 0$ ,  $\delta_4 = -c_1 - c_2$ ,

$$\begin{aligned}\delta_5 &= -\frac{1}{2}(a_1 + a_2 + c_3 + c_4) + \frac{1}{2}\sqrt{(a_2 - a_1 + c_3 + c_4)^2 - 4a_2c_3}, \\ \delta_6 &= -\frac{1}{2}(a_1 + a_2 + c_3 + c_4) - \frac{1}{2}\sqrt{(a_2 - a_1 + c_3 + c_4)^2 - 4a_2c_3}.\end{aligned}$$

By similar argument, we still have that  $\delta_i$  ( $i = 1, 2, \dots, 6$ ) are either zeros or have negative real parts.

## Appendix D.

### Appendix D.1.

In this subsection, we use  $\Re(z)$  and  $\Im(z)$  to respectively represent the real and imaginary parts of a complex number  $z$ . The following is mainly extracted from [51] (page 39~42).

**Theorem 11. (*The Jordan Canonical Form*).** *Let  $A$  be a real matrix with real eigenvalues  $\lambda_j$ ,  $j = 1, \dots, k$  and complex eigenvalues  $\lambda_j = a_j + ib_j$  and  $\bar{\lambda}_j = a_j - ib_j$ ,  $j = k+1, \dots, n$ . Then there exists a basis  $\{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}, \mathbf{u}_{k+1}, \dots, \mathbf{v}_n, \mathbf{u}_n\}$  for  $\mathbb{R}^{2n-k}$  where  $\mathbf{v}_j$ ,  $j = 1, \dots, k$  and  $\mathbf{w}_j$ ,  $j = k+1, \dots, n$  are generalized eigenvectors of  $A$ ,  $\mathbf{u}_j = \Re(\mathbf{w}_j)$  and  $\mathbf{v}_j = \Im(\mathbf{w}_j)$  for  $j = k+1, \dots, n$ , such that the matrix  $P = \{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}, \mathbf{u}_{k+1}, \dots, \mathbf{v}_n, \mathbf{u}_n\}$  is invertible and*

$$P^{-1}AP = \begin{bmatrix} B_1 & & \\ & \ddots & \\ & & B_r \end{bmatrix}$$

where the elementary Jordan blocks  $B = B_j$ ,  $j = 1, \dots, r$  are either of the form

$$B = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \cdots & & & & \\ 0 & \cdots & & \lambda & 1 \\ 0 & \cdots & & 0 & \lambda \end{bmatrix} \quad (\text{D.1})$$

for  $\lambda$  one of the real eigenvalues of  $A$  or of the form

$$B = \begin{bmatrix} D & I_2 & 0 & \cdots & 0 \\ 0 & D & I_2 & \cdots & 0 \\ \cdots & & & & \\ 0 & \cdots & & D & I_2 \\ 0 & \cdots & & 0 & D \end{bmatrix} \quad (\text{D.2})$$

with

$$D = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}, \quad I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad 0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

for  $\lambda = a + ib$  one of the complex eigenvalues of  $A$ .

The Jordan canonical form of  $A$  yields some explicit information about the form of  $\mathbf{x} = e^{At}\mathbf{x}_0$ , i.e. the solution of the initial value problem

$$\begin{cases} \frac{d\mathbf{x}}{dt} = A\mathbf{x} \\ \mathbf{x}(0) = \mathbf{x}_0 \end{cases} \quad (\text{D.3})$$

That is,

$$\mathbf{x}(t) = P \text{diag} [e^{B_j t}] P^{-1} \mathbf{x}_0, \quad (\text{D.4})$$

where  $B_j$  are the elementary Jordan blocks of  $A$ ,  $j = 1, \dots, r$ . Here  $\text{diag} [e^{B_j t}]$  represents

$$\text{diag} [e^{B_j t}] = \begin{pmatrix} e^{B_1 t} & 0 & \cdots & 0 \\ 0 & e^{B_2 t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{B_r t} \end{pmatrix}.$$

In the following, the notation  $\text{diag}[\cdot]$  indicates the similar meaning. If  $B_j = B$  is an  $m \times m$  matrix of the form (D.1) and  $\lambda$  is a real eigenvalue of  $A$  then

$$e^{Bt} = e^{\lambda t} \begin{bmatrix} 1 & t & t^2/2! & \cdots & t^{m-1}/(m-1)! \\ 0 & 1 & t & \cdots & t^{m-2}/(m-2)! \\ 0 & 0 & 1 & \cdots & t^{m-3}/(m-3)! \\ \cdots & & & & \\ 0 & \cdots & & 1 & t \\ 0 & \cdots & & 0 & 1 \end{bmatrix}. \quad (\text{D.5})$$

If  $B_j = B$  is an  $2m \times 2m$  matrix of the form (D.2) and  $\lambda = a + ib$  is a complex eigenvalue of  $A$ , then

$$e^{Bt} = e^{at} \begin{bmatrix} R & Rt & Rt^2/2! & \cdots & Rt^{m-1}/(m-1)! \\ 0 & R & Rt & \cdots & Rt^{m-2}/(m-2)! \\ 0 & 0 & R & \cdots & Rt^{m-3}/(m-3)! \\ \cdots & & & & \\ 0 & \cdots & & R & Rt \\ 0 & \cdots & & 0 & R \end{bmatrix} \quad (\text{D.6})$$

where  $R$  is the rotation matrix

$$R = \begin{bmatrix} \cos bt & -\sin bt \\ \sin bt & \cos bt \end{bmatrix}.$$

**Theorem 12.** *If  $\mathbf{x}(t)$  satisfies (D.3), then each coordinate in  $\mathbf{x}(t)$  is a linear combination of functions of the form*

$$t^k e^{at} \cos bt \quad \text{or} \quad t^k e^{at} \sin bt$$

where  $\lambda = a + ib$  is an eigenvalue of the matrix  $A_{n \times n}$  and  $0 \leq k \leq n - 1$ .

**Corollary 3.** *If the eigenvalues of  $A$  are either zeros or have negative real parts, and  $\mathbf{x}(t)$  is bounded in  $[0, \infty)$ , then  $\mathbf{x}(t)$  converges to a finite limit as time goes to infinity.*

*Proof.* The solution is composed of the terms like  $t^k e^{at} \cos bt$  and  $t^k e^{at} \sin bt$ . If  $a < 0$ , then  $t^k e^{at} \cos bt$  and  $t^k e^{at} \sin bt$  converge as time goes to infinity. If  $a = b = 0$ , we will see  $k = 0$ . In fact, in this case  $t^k e^{at} \cos bt = t^k$ . If  $k > 0$ , then this term  $t^k$  in the solution will make the solution unbounded as  $t$  tends to infinity. So  $k$  must be zero in the terms corresponding to  $a = b = 0$ . Thus,  $t^k e^{at} \cos bt = 1$  and  $t^k e^{at} \sin bt = 0$ . So the solution converges.  $\square$

#### Appendix D.2.

As the above subsections illustrate, the following problem

$$\begin{cases} \frac{d\mathbf{x}}{dt} = A\mathbf{x} \\ \mathbf{x}(0) = \mathbf{x}_0 \end{cases}$$

has solution  $\mathbf{x}(t) = e^{At} \mathbf{x}_0$ , which equals

$$\mathbf{x}(t) = P \operatorname{diag} [e^{B_j t}] P^{-1} \mathbf{x}_0, \quad (\text{D.7})$$

where  $B_j$  are the elementary Jordan blocks of  $A$ ,  $j = 1, \dots, r$ . Suppose the rank of  $A_{n \times n}$  is  $n - 1$ . This means that zero is a one fold eigenvalue of  $A$ .

According to (D.7), we construct a corresponding  $\hat{\mathbf{x}}$  in the form

$$\hat{\mathbf{x}} = P \operatorname{diag} [\hat{B}_j] P^{-1} \mathbf{x}_0, \quad (\text{D.8})$$

where  $\hat{B}_j(t)$  is defined as follows. If  $e^{B_j t}$  in (D.7) has the form of (D.6), then  $\hat{B}_j$  is defined by

$$\hat{B}_j = \mathbf{0}_{2m \times 2m}. \quad (\text{D.9})$$

If  $e^{B_j t}$  has the form of (D.5), and the corresponding real eigenvalue  $\lambda < 0$ , then  $\hat{B}_j$  is defined by

$$\hat{B}_j = \mathbf{0}_{m \times m} \quad (\text{D.10})$$

If  $\lambda = 0$ , we know that zero is a one fold eigenvalue of  $A$  due to its rank  $n - 1$ . So  $m = 1$ . Then,

$$\hat{B}_j = 1. \quad (\text{D.11})$$

In short, only for the zero eigenvalue is  $\hat{B}_j$  set to one, otherwise it is set to zeros. Clearly, we have

**Lemma 16.** *If zero is a one fold eigenvalue of  $A$  and all other eigenvalues of  $A$  have negative real parts, then*

$$\lim_{t \rightarrow \infty} |\hat{\mathbf{x}}(t) - \mathbf{x}(t)| = 0.$$

*Proof.* If  $B$  is the Jordan block corresponding to the one fold zero eigenvalue, then according to (D.11),  $e^{Bt} = e^0 = 1$ , and  $e^{Bt} - \hat{B} = 1 - 1 = 0$ . For any non-zero eigenvalue, since  $\hat{B} = 0$  then  $e^{Bt} - \hat{B} = e^{Bt}$ . Notice that by (D.5) and (D.6),

$$\|e^{Bt}\| \leq C_1(t)e^{-\Lambda t},$$

where  $C_1(t)$  is a polynomial of  $t$  with the maximum order  $k$ , and

$$\Lambda = \inf\{-\Re(\lambda) \mid \lambda \text{ is non-zero eigenvalue of } A\} > 0.$$

Therefore,

$$\|\mathbf{x}(t) - \hat{\mathbf{x}}\| \leq \sum_B \|e^{Bt} - \hat{B}\| \leq C_2(t)e^{-\Lambda t} \longrightarrow 0 \quad (\text{D.12})$$

as  $t$  goes to infinity, where  $C_2(t)$  is a polynomial function of  $t$ .  $\square$

The above construction can be extended to the problem of

$$\frac{dX}{dt} = A(t)X \quad (\text{D.13})$$

with the initial value  $X(0)$ , which is discussed in Section 7.2.3. The solution is

$$X(t) = e^{tC(t)} X(0),$$

where  $C(t) = \frac{1}{t} \int_0^t A(s) ds$ . We can similarly define a function  $f$  such that  $C(t) = f(\beta(t))$ , where  $\beta(t)$  is similarly defined according to  $A(t)$ . Therefore,

$$X(t) = e^{tC(t)} X(0) = e^{tf(\beta(t))} X(0).$$

For a fixed  $\beta$ ,

$$e^{tf(\beta)} X(0) = P(\beta) \text{diag} [e^{B_j(\beta)t}] P(\beta)^{-1} X(0), \quad (\text{D.14})$$

where  $B_j(\beta)$  are the elementary Jordan blocks of  $f(\beta)$ ,  $j = 1, \dots, r(\beta)$ . Repeating the previous construction process with  $B_j(\beta)$  for each  $j$ , we obtain the constructed matrix  $\hat{B}(\beta)_j$ . We define

$$h(\beta) = P(\beta) \text{diag} [\hat{B}(\beta)_j] P(\beta)^{-1} X(0). \quad (\text{D.15})$$

For convenience, suppose the dimension of  $A(t)$  in (D.13) is  $n \times n$ . We should point out that for any  $t$ , the rank of  $A(t)$  is  $n - 1$ , and thus for any  $t$ ,  $A(t)$ 's zero eigenvalue is one fold. In fact, the rank of any infinitesimal generator with dimension  $n \times n$  is  $n - 1$ . This implies that any  $n - 1$  columns or rows of this generator are linearly independent. According to the definition of  $A(t)$  in Section 7.2.3,  $A(t)$  is an infinitesimal generator if  $\alpha(t) \neq 0$ . If  $\alpha(t) = 0$ ,  $A(t)$  is also a generator after one column is modified (see Proposition 14 and 15), which means that the other  $n - 1$  columns are linearly independent. So whatever  $t$  is, the rank of  $A(t)$  is  $n - 1$ . Thus, the zero eigenvalue is one fold. Therefore,  $f(\beta)$ 's zero eigenvalue is also one fold for any  $\beta$ . So each entry of all blocks  $\hat{B}(\beta)_j$  is zero, except for the one corresponding to the zero eigenvalue, in which case this block is a scalar one. This implies that for any  $\beta$  all entries of the matrix  $\text{diag}[\hat{B}(\beta)_j]$  are zeros, except for a diagonal entry with one.

By permutation,  $\text{diag} [\hat{B}(\beta)_j]$  can always be transformed into the form

$$\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Correspondingly,  $P(\beta)$  is permuted into  $U(\beta)$ . Therefore, the formulae (D.15) can be written as

$$h(\beta) = U(\beta) \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} U(\beta)^{-1} X(0), \quad (\text{D.16})$$

Now we prove a proposition which is used in Section 7.2.3.

**Proposition 16.** *For the  $e^{tf(\beta)} X(0)$  in (D.14) and  $h(\beta)$  in (D.15), we have*

$$\lim_{t \rightarrow \infty} \|e^{tf(\beta)} X(0) - h(\beta)\| = 0.$$

*Proof.* By a similar estimation as (D.12), we have

$$\|e^{tf(\beta)} X(0) - h(\beta)\| \leq C(t)e^{-\Lambda_1 t}.$$

where  $C(t)$  is a polynomial of  $t$ . By a similar proof to Lemma 8, we have

$$\Lambda_1 = \inf_{\beta \in [0,1]} \{-\Re(\lambda) | \lambda \text{ is } f(\beta)\text{'s non-zero eigenvalue}\} > 0.$$

Then

$$\|e^{tf(\beta)} X(0) - h(\beta)\| \leq C(t)e^{-\Lambda_1 t} \longrightarrow 0$$

as  $t$  goes to infinity. □

Let  $\hat{X}(t) = h(\beta(t))$ , where  $\beta(t) \in [0, 1]$ , and notice  $X(t) = f(\beta(t))$ . As a consequence of this proposition, we have

**Corollary 4.** *Let  $X(t)$  be the solution of  $\frac{dX}{dt} = A(t)X$  which is discussed in Section 7.2.3, and let  $\hat{X}(t) = h(\beta(t))$ , then*

$$\lim_{t \rightarrow \infty} \|X(t) - \hat{X}(t)\| = 0.$$

*Appendix D.3.*

This subsection presents a proof of Lemma 7 in Section 7.1.3. Let  $\lambda(\beta)$  be a nonzero eigenvalue of the following  $f(\beta)$ :

$$f(\beta) = \begin{pmatrix} -a\beta & b \\ a\beta & -b \end{pmatrix}.$$

We will prove the following lemma which states that the real part of  $\lambda(\beta)$  is negative. The proof given here does not rely on the explicit expression of the eigenvalue.

**Lemma 7:** For any  $\beta \in [0, 1]$ ,  $\Re(\lambda(\beta)) < 0$ , where  $\lambda(\beta)$  is a nonzero eigenvalue of  $f(\beta)$ .

*Proof.* After a shift  $\max\{a\beta, b\}I$ ,  $f(\beta)$  becomes  $\tilde{f}(\beta) = f(\beta) + \max\{a\beta, b\}I$ , which is a nonnegative matrix. Then similarly to the proof of Theorem 7, which is based on the Perron-Frobenius theorem (Theorem 6), we can conclude that the eigenvalue other than zero has negative real part.  $\square$